INVITED PAPER

Image Reconstruction: From Sparsity to Data-Adaptive Methods and Machine Learning

This article overviews how sparsity, data-driven methods and machine learning have, and will continue to, influence the general area of image reconstruction, cutting across modalities. In general, this contribution looks at progress in medical image reconstruction methods with focus on the two most recent trends: methods based on sparsity or low-rank models, and data-driven methods based on machine learning techniques.

By Saiprasad Ravishankar, *Member IEEE*, Jong Chul Ye, *Senior Member IEEE*, and Jeffrey A. Fessler, *Fellow IEEE*

ABSTRACT | The field of medical image reconstruction has seen roughly four types of methods. The first type tended to be analytical methods, such as filtered backprojection (FBP) for X-ray computed tomography (CT) and the inverse Fourier transform for magnetic resonance imaging (MRI), based on simple mathematical models for the imaging systems. These methods are typically fast, but have suboptimal properties such as poor resolution-noise tradeoff for CT. A second type is iterative reconstruction methods based on more complete models for the imaging system physics and, where appropriate, models for the sensor statistics. These iterative methods improved image quality by reducing noise and artifacts. The U.S. Food and Drug Administration (FDA)-approved methods among these have been based on relatively simple regularization models. A third type of methods has been designed to accommodate modified data acquisition methods, such as reduced sampling in MRI and CT to reduce scan time or radiation dose. These methods typically involve mathematical image models involving assumptions such as *sparsity* or *low rank*. A fourth type of methods replaces mathematically designed models of signals and systems with *data-driven* or *adaptive* models inspired by the field of *machine learning*. This article focuses on the two most recent trends in medical image reconstruction: methods based on sparsity or low-rank models and data-driven methods based on machine learning techniques.

KEYWORDS | Compressed sensing (CS); deep learning; dictionary learning (DL); efficient algorithms; image reconstruction; machine learning; magnetic resonance imaging (MRI); multilayer models; nonconvex optimization; positron emission tomography (PET); single-photon emission computed tomography (SPECT); sparse and low-rank models; structured models; transform learning; X-ray computed tomography (CT).

Manuscript received April 4, 2019; revised June 25, 2019 and August 6, 2019; accepted August 7, 2019. Date of publication September 19, 2019; date of current version December 26, 2019. The work of J. C. Ye was supported in part by the National Research Foundation of Korea under Grant 2016R1A2B3008104. The work of J. A. Fessler was supported in part by the National Institutes of Health (NIH) under Grant R01 CA214981, Grant R01 EB023618, Grant U01 EB018753, and Grant U01 EB026977. *(Corresponding author: Saiprasad Ravishankar.)*

S. Ravishankar is with the Departments of Computational Mathematics, Science and Engineering, and Biomedical Engineering, Michigan State University, East Lansing, MI 48824 USA (e-mail: ravisha3@msu.edu).

J. C. Ye is with the Department of Bio and Brain Engineering, Department of Mathematical Sciences, Korea Advanced Institute of Science and Technology (KAIST), Daejeon 34141, South Korea (e-mail: jong.ye@kaist.ac.kr).

J. A. Fessler is with the Department of Electrical Engineering and Computer Science, University of Michigan, Ann Arbor, MI 48109 USA (e-mail: fessler@umich.edu).

Digital Object Identifier 10.1109/JPROC.2019.2936204

I. INTRODUCTION

Various medical imaging modalities are popular in clinical practice, such as magnetic resonance imaging (MRI), X-ray computed tomography (CT), positron-emission tomography (PET), single-photon emission CT (SPECT), and so

on. These modalities help to image various biological and anatomical structures and physiological functions and aid in medical diagnosis and treatment. Ensuring high quality images reconstructed from limited or corrupted (e.g., noisy) measurements such as subsampled data in MRI (reducing acquisition time) or low-dose or sparse-view data in CT (reducing patient radiation exposure) has been a popular area of research and holds high value in improving clinical throughput and patient experience. This article reviews some of the major recent advances in the field of image reconstruction, focusing on methods that use sparsity, low-rankness, and machine learning. We focus partly on PET, SPECT, CT, and MRI examples, but the general methods can be useful for other modalities, both medical and nonmedical. This article is part of a special issue that focuses on sparsity and machine learning in medical imaging. Other articles in this issue emphasize other modalities.

A. Types of Image Reconstruction Methods

Image reconstruction methods have undergone significant advances over the past few decades, with different paths for various modalities. These advances can be broadly grouped in four categories of methods. The first category consists of analytical and algebraic methods. These methods include the classical filtered backprojection (FBP) methods for X-ray CT [e.g., Feldkamp–Davis–Kress (FDK) method [1]] and the inverse Fast Fourier transform and extensions such as the nonuniform fast Fourier transform (NUFFT) [2], [3] for MRI and CT. These methods are based on relatively simple mathematical models of the imaging systems, and although they have efficient and fast implementations, they suffer from suboptimal properties such as poor resolution-noise tradeoff for CT.

A second category of reconstruction methods involves iterative reconstruction algorithms that are based on more sophisticated models for the imaging system's physics and models for sensor and noise statistics. Often called model-based image reconstruction (MBIR) methods or statistical image reconstruction (SIR) methods, these schemes iteratively estimate the unknown image based on the system (physical or forward) model, measurement statistical model, and assumed prior information about the underlying object [4], [5]. For example, minimizing penalized weighted-least-squares (PWLS) cost functions has been popular in many modalities including PET and X-ray CT, and these costs include a statistically weighted quadratic data-fidelity term (capturing the imaging forward model and noise variance) and a penalty term called a regularizer that models the prior information about the object [6]. These iterative reconstruction methods improve image quality by reducing noise and artifacts. In MRI, parallel data acquisition methods (P-MRI) exploit the diversity of multiple receiver coils to acquire fewer Fourier or k-space samples [7]. Today, P-MRI acquisition is used widely in commercial systems, and MBIR-type methods in this case include those based on coil sensitivity encoding (SENSE) [7], and so on. The iterative medical image reconstruction methods approved by the U.S. Food and Drug Administration (FDA) for SPECT, PET, and X-ray CT have been based on relatively simple regularization models.

A third category of reconstruction methods accommodate modified data acquisition methods such as reduced sampling in MRI and CT to significantly reduce scan time and/or radiation dose. Compressed sensing (CS) techniques [8]–[12] have been particularly popular among this class of methods (leading to journal special issues [13], [14]). These methods have been so beneficial for MRI [15], [16] that they recently obtained FDA approval [17]–[19]. CS theory predicts the recovery of images from far fewer measurements than the number of unknowns, provided that the image is sparse in a transform domain or dictionary, and the acquisition or sampling procedure is appropriately incoherent with the transform. Since magnetic resonance (MR) acquisition in Fourier or k-space occurs sequentially over time, making it a relatively slow modality, CS for MRI can enable quicker acquisition by collecting fewer k-space samples. However, the reduced sampling time comes at the cost of slower, nonlinear, and iterative reconstruction. The methods for reconstruction from limited data typically exploit mathematical image models based on *sparsity* or *low rank*, and so on. In particular, CS-based MRI methods often use variable density random sampling techniques to acquire the data and use sparsifying transforms such as wavelets, finite difference operators [via total variation (TV) penalty], contourlets, etc., for reconstruction [15], [20]. Research about such methods also focused on developing new theory and guarantees for sampling and reconstruction from limited data [21] and on new optimization algorithms for reconstruction with good convergence rates [22].

A fourth category of image reconstruction methods replaces mathematically designed models of images and processes with *data-driven* or *adaptive* models inspired by the field of *machine learning*. Such models (e.g., synthesis dictionaries [23], sparsifying transforms [24], tensor models, etc.) can be learned in various ways such as by using training data sets [25], [26], or even learned jointly with the reconstruction [25], [27]–[29], a setting called model-blind reconstruction or blind CS (BCS) [30]. Although most of these methods perform offline reconstruction (where the reconstruction is performed once all the measurements are collected), recent works show that the models can also be learned in a time-sequential or online manner from streaming measurements to reconstruct dynamic objects [31], [32]. The learning can be done in an unsupervised manner employing model-based and surrogate cost functions, or the reconstruction algorithms [such as deep convolutional neural networks (CNNs)] can be trained in a supervised manner to minimize the error in reconstructing training data sets that typically consist of pairs of ground truth and undersampled data [33]–[37]. These learning-based reconstruction

methods form a very active field of research with numerous conference special sessions and special journal issues devoted to the topic [38].

The categories above are not a strict chronology; for example, neural network (NN) methods were investigated for image reconstruction as early as 1991 [39], and for MR spectroscopy soon thereafter [40], and some of the earliest methods for X-ray CT were iterative.

B. Focus and Outline of This Article

This article reviews the progress in medical image reconstruction, focusing on the two most recent trends: methods based on sparsity using analytical models, and low-rank models and extensions that combine sparsity and low rank, etc.; and data-driven models and approaches exploiting machine learning. Some of the mathematical underpinnings and connections between different models and their pros and cons are also discussed.

This article is organized as follows. Section II describes early image reconstruction approaches, especially those used in current clinical systems. Sections III and IV describe sparsity and low-rank-based approaches for image reconstruction. Section V surveys the advances in data-driven image models and related machine learning approaches for image reconstruction. Among the learning-based methods, techniques that learn image models using model-based cost functions from the training data, or on-the-fly from measurements are discussed, followed by recent methods relying on supervised learning of models for reconstruction, typically from data sets of high-quality images and their corrupted versions. Section VI reviews the very recent works using learned CNNs (also known as deep learning) for image reconstruction. Section VII discusses some of the current challenges and open questions in image reconstruction and outlines future directions for the field. Section VIII concludes this review article.

II. ITERATIVE RECONSTRUCTION USED CLINICALLY

This section focuses on some of the iterative MBIR methods that are in routine clinical use currently and relates the models used in those systems to the sparsity models used in the contemporary literature. As mentioned in the introduction, MBIR methods have been used routinely for many years in commercial SPECT, PET, and CT systems. Early publications on MBIR methods tended to focus on mathematical Bayesian models. In contrast, recent data-driven methods are based on empirical distributions from the training data, as discussed later in this article. The dominant Bayesian approach for reconstructing an image $\hat{\boldsymbol{x}}$ from data $\boldsymbol{y}$ was the maximum *a posteriori* (MAP) approach of finding the maximizer of the posterior $p(\boldsymbol{x}|\boldsymbol{y})$. By the Bayes rule, the MAP approach is equivalent to

$$\hat{\boldsymbol{x}} = \arg\min_{\boldsymbol{x}} \left(-\log p(\boldsymbol{y}|\boldsymbol{x}) - \log p(\boldsymbol{x}) \right) \tag{1}$$

where $-\log p(\boldsymbol{y}|\boldsymbol{x})$ denotes the negative log-likelihood that describes the imaging system physics and noise statistics. The benefits of modeling the system noise and physics properties were the primary driver for the early work on MBIR methods for PET and SPECT, compared to classical reconstruction methods like FBP that use quite simple geometric models and lack statistical modeling. In MRI, early iterative methods were driven by non-Cartesian sampling and parallel imaging [41]. The function $p(\boldsymbol{x})$ in (1) denotes a Bayesian prior that captures assumptions about the image $\boldsymbol{x}$. Markov random field (MRF) models were particularly popular in early work; these methods typically assign higher prior probabilities for images $\boldsymbol{x}$ where neighboring pixels tend to have similar values [42], [43], often using "line sites" to infer the presence of boundaries between pixels [44], sometimes with the guidance of images from other modalities of the same patient ("anatomical priors") [45]–[48].

Although the term *sparsity* is uncommon in papers about MRF models, the "older" assumption that neighboring pixels tend to have similar values is quite closely related to the "newer" assumption that the differences between neighboring pixel values tend to be sparse.

The form of (1) is equivalent[1] to the following regularized optimization problem:

$$\hat{\boldsymbol{x}} = \arg\min_{\boldsymbol{x}} f(\boldsymbol{x}) + \beta R(\boldsymbol{x}) \tag{2}$$

where $f(\boldsymbol{x})$ denotes a data-fidelity term and $R(\boldsymbol{x})$ denotes a regularizer that encourages the image $\hat{\boldsymbol{x}}$ to have some assumed properties such as piecewise smoothness. The positive regularization parameter β controls the tradeoff between overfitting the (noisy) data and oversmoothing the image. More recent MBIR papers, and the commercial methods, tend to adopt this regularization perspective rather than using the Bayesian terminology. Early commercial PET and SPECT reconstruction methods used unregularized algorithms [49], but more recent methods use edge-preserving (EP) regularization involving nonquadratic functions of the differences between neighboring pixels [50], essentially implicitly assuming that the image gradients are sparse, i.e., that those differences are mostly zero or near zero. In 1-D, a typical regularizer would be

$$R(\boldsymbol{x}) = \textstyle\sum_{n=2}^{N} \psi(x_n - x_{n-1}) \tag{3}$$

where N is the number of pixels, and ψ denotes a "potential function" (in Bayesian parlance) such as the hyperbola $\psi(z) = (|z|^2 + \delta^2)^{1/2}$ or a generalized Gaussian function [51] (a few modifications of the regularizer

[1] For any prior $p(\boldsymbol{x})$, one can simply define $\beta R(\boldsymbol{x}) = -\log p(\boldsymbol{x})$ to write (1) in the form (2). However, for most regularizers $R(\boldsymbol{x})$ used in practice, defining $p(\boldsymbol{x}) \propto \exp(-\beta \boldsymbol{R}(\boldsymbol{x}))$ would be an "improper prior" because there is no constant that makes $p(\boldsymbol{x})$ integrate to unity.

are needed to make it work well in practice [52], [53]). MBIR methods for clinical CT systems also use EP regularization [54].

The potential functions ψ that are used clinically in FDA-approved methods for CT and PET include a generalized Gaussian function [54] and a relative-difference prior [53]. These potential functions are relatives of the total variation (TV) regularizer that is studied widely in the academic literature. However, TV imposes a strong assumption of gradient sparsity because it uses the nonsmooth absolute value potential $\psi(z) = |z|$ that is well suited to images that are piecewise *constant* but less suitable for images that are piecewise *smooth*. In particular, the TV regularizer leads to CT images with undesirable patchy textures; so the commercial systems use an EP regularizer that does not enforce sparsity as strictly [54]. In summary, the current clinical methods for PET, SPECT, and CT use optimization formulations of the form (2) with regularizers akin to (3), thereby moderately encouraging gradient sparsity.

III. SPARSITY USING MATHEMATICAL MODELS

This section discusses image reconstruction methods that are based on models for the image $\boldsymbol{x}$ that involve some form of sparsity. Methods based on sparsity models have a long history in signal processing [55], [56]. Such methods are now being used clinically to accelerate MRI scans, making such scans shorter, reducing the effects of patient motion, and improving patient comfort.

The regularizer based on finite differences in (3) (e.g., with $\psi(z) = |z|$) is equivalent to assuming the image gradients are sparse. Assumptions of gradient sparsity or piecewise smoothness have a long history in imaging [42], [43], [57]. This model is a special case of the more general assumption that $\boldsymbol{T}\boldsymbol{x}$ is sparse for some spatial operator $\boldsymbol{T}$. This is called "*analysis regularization*" and a typical image reconstruction optimization formulation for such models is

$$\hat{\boldsymbol{x}} = \arg\min_{\boldsymbol{x}} \frac{1}{2} \|\boldsymbol{A}\boldsymbol{x} - \boldsymbol{y}\|_2^2 + \beta \|\boldsymbol{T}\boldsymbol{x}\|_1 . \tag{4}$$

The ℓ_1-norm is often used as the sparsity regularizer and can be viewed as a convex relaxation or convex envelope of the nonconvex ℓ_0 "norm" that measures the size of the support of a vector or counts the number of nonzero entries. Alternative penalties such as $\|\boldsymbol{T}\boldsymbol{x}\|_p^p$ for $0 < p < 1$ that better approximate $\|\boldsymbol{T}\boldsymbol{x}\|_0$ have also been used for reconstruction [58]. There are many operators $\boldsymbol{T}$ that have been used for image reconstruction; the two most popular ones are finite-differences, corresponding to TV, and various wavelet transforms. Wavelets are the sparsifying model used in the JPEG 2000 image compression standard, because they are effective at sparsifying natural images. The combination of both wavelets and TV is particularly common in MRI [15], and, although the details are proprietary, it is likely that such combinations are used in the commercial MRI systems [59].

In some settings, one has a "prior image" $\bar{\boldsymbol{x}}$ available, in which case one can modify (4) to encourage similarity with that prior image using a cost function like

$$\hat{\boldsymbol{x}} = \arg\min_{\boldsymbol{x}} \frac{1}{2} \|\boldsymbol{A}\boldsymbol{x} - \boldsymbol{y}\|_2^2 + \beta \|\boldsymbol{T}(\boldsymbol{x} - \bar{\boldsymbol{x}})\|_1 . \tag{5}$$

The prior image constrained CS (PICCS) approach is an example of this type of approach [60].

An alternative to the analysis regularization model (4) is to assume that the image can be represented as a sparse linear combination of atoms from a dictionary, i.e., $\boldsymbol{x} = \boldsymbol{D}\boldsymbol{z}$ where $\boldsymbol{D}$ is a dictionary and $\boldsymbol{z}$ is a coefficient vector. One way to express this assumption as an optimization problem is

$$\hat{\boldsymbol{x}} = \boldsymbol{D}\hat{\boldsymbol{z}}, \quad \hat{\boldsymbol{z}} = \arg\min_{\boldsymbol{z}} \frac{1}{2} \|\boldsymbol{A}\boldsymbol{D}\boldsymbol{z} - \boldsymbol{y}\|_2^2 + \beta \|\boldsymbol{z}\|_1 \tag{6}$$

where $\boldsymbol{A}$ denotes the imaging system model. This synthesis formulation is equivalent to the analysis formulation (4) when $\boldsymbol{D}$ and $\boldsymbol{T}$ are both square and full-rank and $\boldsymbol{D} = \boldsymbol{T}^{-1}$ (a basis). When $\boldsymbol{T}$ is tall and its rows form a frame, then (4) is equivalent to a synthesis formulation ($\boldsymbol{D} = \boldsymbol{T}^L$, a left-inverse of $\boldsymbol{T}$), but with $\boldsymbol{z}$ in (6) constrained to be in the range space of $\boldsymbol{T}$. However, usually, $\boldsymbol{T}$ is a general operator and $\boldsymbol{D}$ is wide. A drawback of this synthesis sparsity formulation is that it relies heavily on the assumption that $\boldsymbol{x} = \boldsymbol{D}\boldsymbol{z}$, whereas an approximate form $\boldsymbol{x} \approx \boldsymbol{D}\boldsymbol{z}$ may be more reasonable in practice, particularly when the dictionary $\boldsymbol{D}$ comes from a mathematical model that might not perfectly represent natural medical images. An alternative synthesis formulation that allows an approximate sparsity model is

$$\begin{aligned} \hat{\boldsymbol{x}} &= \arg\min_{\boldsymbol{x}} \frac{1}{2} \|\boldsymbol{A}\boldsymbol{x} - \boldsymbol{y}\|_2^2 + \beta R(\boldsymbol{x}) \\ R(\boldsymbol{x}) &= \min_{\boldsymbol{z}} \frac{1}{2} \|\boldsymbol{x} - \boldsymbol{D}\boldsymbol{z}\|_2^2 + \alpha \|\boldsymbol{z}\|_1 . \end{aligned}$$

A drawback of this approach is that it requires one to select two regularization parameters (α and β). Other sparsity models include generalized analysis models [24], and the balanced sparse model for tight frames [61], where the signal is sparse in a synthesis dictionary and also approximately sparse in the corresponding transform (transpose of the dictionary) domain, with a common sparse representation in both domains. These models have been applied to inverse problems such as in CS MRI [15], [61], [62].

The drawback of all of the models discussed in this section is that, traditionally, the underlying operators such as $\boldsymbol{T}$ and $\boldsymbol{D}$ are designed mathematically, typically with empirical validation on real data, rather than being computed directly from training data or adapted to a specific

patient's data. Nevertheless, they are useful, as evidenced by their adoption in clinical MRI systems. Most of the methods in subsequent sections are more data-driven approaches.

IV. LOW-RANK MODELS

Although sparsity models have been popular in image reconstruction, particularly in CS, various alternative models exploiting properties such as the inherent low rankness of the data have also shown promise in imaging applications. This section reviews some of the low-rank models and their extensions such as when combined with sparsity, followed by recent structured low-rank matrix approaches [63]–[68]. The assumption that a matrix is low rank is equivalent to assuming that its singular values are sparse. Both sparsity and low rankness involve assumptions of simplicity, and these relationships are unified by the notion of atomic norms [69]. As there are many works on low-rank models for various modalities and applications, we only provide an overview of some low-rank methods, their extensions, and connections to sparsity, rather than an exhaustive review.

A. Low-Rank Models and Extensions

Low-rank models have been exploited in many imaging applications such as dynamic MRI [70], functional MRI [71], [72], diffusion-weighted MRI [73], and MR fingerprinting [74]–[76].

Low-rank assumptions are especially useful when processing dynamic or time-series data and have been popular in dynamic MRI, where the underlying image sequence tends to be quite correlated over time. In dynamic MRI, the measurements are inherently undersampled because the object changes as the samples are collected. Reconstruction methods therefore typically pool the k-t space data in time to make sets of k-space data (the underlying dynamic object is written in the form of a Casorati matrix [70], whose rows represent voxels and columns denote temporal frames, and the sets of k-space data denote measurements of such frames) that appear to have sufficient samples. However, these methods can have poor temporal resolution and artifacts due to pooling. Careful design of model-based (CS-type) techniques can help to achieve improved temporal or spatial resolution in such undersampled settings.

Several works have exploited low rankness of the underlying Casorati (space-time) matrix for dynamic MRI reconstruction [70], [77]–[79]. Low-rank modeling of local space-time image patches has also been investigated in [80]. Later works combined low-rank (L) and sparsity (S) models for improved reconstruction. Some of these works model the dynamic image sequence as both low-rank and sparse (L & S) [81], [82]. There has also been growing interest in models that decompose the dynamic image sequence into the sum of a low-rank and sparse component [also known as robust principal component analysis (RPCA)] [83], [84]. In this L + S model, the low-rank component can capture the background or slowly changing parts of the dynamic object, whereas the sparse component can capture the dynamics in the foreground such as local motion or contrast changes, and so on.

Recent works have applied the L + S model to dynamic MRI reconstruction [85], [86], with the S component modeled as sparse by itself or in a known transform domain. Accurate reconstructions can be obtained [85] when the underlying L and S components are incoherent (distinguishable) and the k-t space acquisition is appropriately incoherent with these components. The L + S reconstruction problem can be formulated as follows:

$$\min_{\boldsymbol{x}_L, \boldsymbol{x}_S} \frac{1}{2} \left\| \boldsymbol{A}(\boldsymbol{x}_L + \boldsymbol{x}_S) - \boldsymbol{y} \right\|_2^2 + \lambda_L \left\| \boldsymbol{R}_1(\boldsymbol{x}_L) \right\|_* + \lambda_S \left\| \boldsymbol{T}\boldsymbol{x}_S \right\|_1 . \tag{7}$$

Here, the underlying vectorized object satisfies the L + S decomposition $\boldsymbol{x} = \boldsymbol{x}_L + \boldsymbol{x}_S$. The sensing operator $\boldsymbol{A}$ acting on it can take various forms. For example, in parallel imaging of a dynamic object, $\boldsymbol{A}$ performs frame-by-frame multiplication by coil sensitivities (in the SENSE approach) followed by undersampled Fourier encoding. The low-rank regularization penalizes the nuclear norm of $\boldsymbol{R}_1(\boldsymbol{x}_L)$, where $\boldsymbol{R}_1(\cdot)$ reshapes its input into a space-time matrix. The nuclear norm serves as a convex surrogate or envelope for the nonconvex matrix rank. The sparsity penalty on $\boldsymbol{x}_S$ has a similar form as in CS approaches, and λ_L and λ_S are nonnegative weights above. Problem (7) is convex and can be solved using various iterative techniques. Otazo *et al.* [85] used the proximal gradient method, wherein the updates involved simple singular value thresholding (SVT) for the L component and soft thresholding for the S component. Later, we mention a data-driven version of the L + S model in Section V-B. Although the above works used low-rank models of matrices (e.g., obtained by reshaping the underlying multidimensional dynamic object into a space-time matrix), other recent works also used low-rank tensor models of the underlying object (a tensor) in reconstruction [87]–[90].

B. Low-Rank Structured Matrix Models

So far, we discussed low-rank and sparsity models that both involve dimensionality reduction. The former involves a low-dimensional subspace, whereas the latter is typically viewed as unions of such subspaces. This section reviews the structured low-rank methods and elaborates on the connections between sparsity and low-rank modeling.

The low-rank Hankel structure matrix approaches [63]–[68], [91]–[93] have been studied extensively for various imaging problems. Unlike the standard low-rank approaches that are based on the redundancies between similar data, the low-rank Hankel structure matrix approaches are based on the fundamental duality between spatial-domain sparsity and the spectral-domain

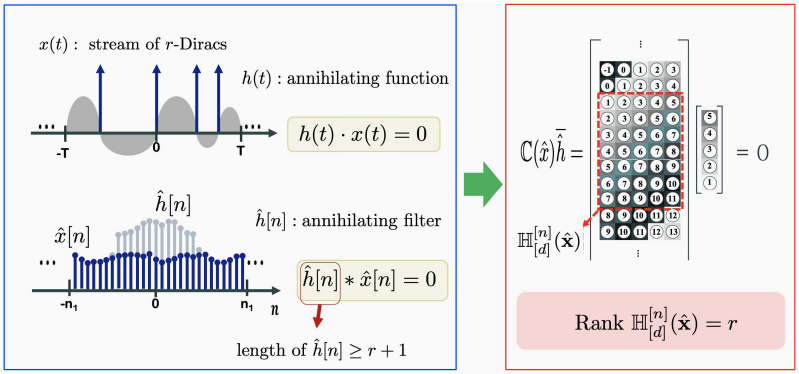

Fig. 1. ***Fundamental duality between sparsity in the image-domain and low-rank Hankel matrix in Fourier domain.***

Hankel matrix rank, which is also related to local polynomial approximation [57].

To explain this duality, we first review the literature on the sampling theory of signals having finite rate of innovations (FRI) [94]–[96]. Consider a superposition of r Dirac impulses as shown in Fig. 1

$$x(t) = \sum_{j=0}^{r-1} a_j \delta\left(t - t_j\right), \quad t_j \in [0, 1]. \tag{8}$$

The associated Fourier series coefficients are given by

$$\hat{x}[k] = \sum_{j=0}^{r-1} a_j e^{-i2\pi k t_j}. \tag{9}$$

The sampling theory for FRI signals [94], [95] showed that there exists an annihilating filter $\hat{h}[k]$ in the Fourier domain, of length $(r+1)$, such that

$$(\hat{h} * \hat{x})[k] = \sum_{l=0}^{r} \hat{h}[l]\hat{x}[k-l] = 0 \tag{10}$$

whose z-transform representation is given by

$$\hat{h}(z) = \sum_{l=0}^{r} \hat{h}[l] z^{-l} = \prod_{j=0}^{r-1} (1 - e^{-i2\pi t_j} z^{-1}). \tag{11}$$

As shown in Fig. 1, the annihilating filter relationship implies that the convolution matrix multiplied by an annihilating filter vector vanishes. Accordingly, the following Hankel structured matrix, corresponding to the submatrix of the convolution matrix, is rank-deficient:

$$\mathbb{H}^{[n]}_{[d]}(\hat{\boldsymbol{x}}) := \begin{bmatrix} \hat{x}[0] & \hat{x}[1] & \cdots & \hat{x}[d-1] \\ \hat{x}[1] & \hat{x}[2] & \cdots & \hat{x}[d] \\ \vdots & \vdots & \ddots & \vdots \\ \hat{x}[n-d] & \hat{x}[n-d+1] & \cdots & \hat{x}[n-1] \end{bmatrix}$$

where $[n] := \{0, \ldots, n-1\}$. Specifically, it was shown in [92] that if the minimum annihilating filter length is $r+1$, then

$$\operatorname{rank} \mathbb{H}^{[n]}_{[d]}(\hat{\boldsymbol{x}}) = r.$$

Thus, given sparsely sampled spectral measurements on the index set $\Omega \subset [n]$, the missing spectrum estimation problem can be formulated as

$$\underset{\boldsymbol{m} \in \mathbb{C}^n}{\arg\min} \ \left\| \mathbb{H}^{[n]}_{[d]}(\boldsymbol{m}) \right\|_* \tag{12}$$

$$\text{s.t. } P_\Omega(\boldsymbol{m}) = P_\Omega(\hat{\boldsymbol{x}}) \tag{13}$$

where $P_\Omega(\cdot)$ denotes the projection on the measured k-space samples on the index set Ω. Although the above discussion is for Dirac impulses, the same principle holds for general FRI signals that can be converted to Diracs or differentiated Diracs after a whitening operator, since the corresponding Fourier spectrum is a simple elementwise multiplication with the spectrums of the operator and the unknown signal, and the weighted spectrum has a low-rank Hankel structure [64], [92].

In contrast to standard CS approaches for MRI, the optimization problem in (12) is purely in the measurement domain. After estimating the fully sampled Fourier data, the final reconstruction can be obtained by a simple inverse Fourier transform. This property leads to remarkable flexibility in real-world applications that classical approaches have difficulty exploiting. For example, this formulation has been successfully applied to CS MRI with state-of-the-art performance for single-coil imaging [63]–[66], [92]. Another flexibility is the recovery of images from multichannel measurements with unknown sensitivities [64], [97]. These schemes rely on the low-rank structure of a structured matrix, obtained by concatenating block Hankel matrices formed from each channel's data. Similar to L + S decomposition in [85], the L + S model for Hankel structure matrix was also used to remove the k-space outliers in MR imaging problems [68]. Such approaches have also been successfully used for super-resolution microscopy [98], image inpainting problems [99], image impulse noise removal [100], and so on.

The common thread between sparsity models, low-rank, and structured low-rank models is that they all strive to

capture signal redundancies to make up for missing or noisy data.

V. DATA-DRIVEN AND LEARNING-BASED MODELS

The most recent class of methods constituting the fourth category of image reconstruction schemes exploits data-driven and learning-based models. This section and the next review several of these varied models and methods. The development of efficient algorithms for the often nonconvex learning-based problems is briefly discussed. The pros and cons of various methods, as well as their connections are also discussed.

A. Partially Data-Adaptive Sparsity-Based Methods

Although early reconstruction methods such as in CS MRI used sparsity in known transform domains such as wavelets [15], TV domain, contourlets [20], etc., later works proposed partially data-adaptive sparsity models by incorporating directional information of patches or block matching, etc., during reconstruction.

A patch-based directional wavelet (PBDW) scheme was proposed for MRI in [101], wherein the regularizer was based on analysis sparsity and was the sum of the ℓ_1-norms of each optimally (adaptively) rearranged and transformed (by fixed 1-D Haar wavelets) image patch. The patch rearrangement or permutation involved rearranging pixels parallel to a certain geometric direction, approximating patch rotation. The best permutation for each patch from among a set of predefined permutations was precomputed based on initial reconstructions to minimize the residual between the transformed permuted patch and its thresholded version. An improved reconstruction method was proposed in [102], where the optimal permutations were computed for patches extracted from the subbands in the 2-D wavelet domain (a shift-invariant discrete wavelet transform is used) of the image. A recent work [103] proposed a different effective modification of the PBDW scheme, wherein a unitary matrix is adapted to sparsify the patches grouped with a common (optimal) permutation. In this case, the analysis sparsity penalty during reconstruction used the ℓ_1-norms of patches transformed by the adapted unitary matrices (one per group of patches).

A different fast and effective method [patch-based nonlocal operator (PANO)] was proposed in [104], wherein for each patch, a small group of patches most similar to it was preestimated (called *block matching*), and the regularizer during reconstruction penalized the sparsity of the groups of patches in a known transform domain. Another reconstruction scheme based on adaptive clustering of patches was applied to MRI in [105]. All these aforementioned methods are also quite related to the recent transform learning-based methods described in Section V-C, where the sparsifying operators are fully adapted in an optimization framework.

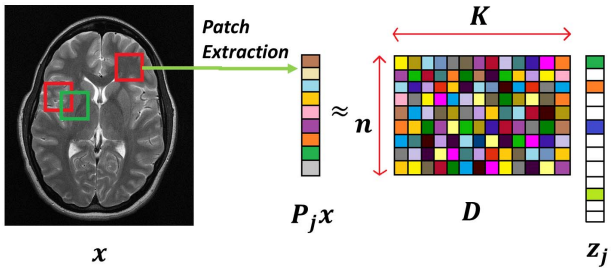

Fig. 2. *Synthesis dictionary model for image patches: overlapping patches $P_j x$ of the image x are assumed approximated by sparse linear combinations of columns of the dictionary D, i.e., $P_j x \approx D z_j$, where z_j has several zeros (denoted with white blocks).*

B. Synthesis Dictionary Learning-Based Approaches for Reconstruction

Among the learning-based approaches that have shown promise for medical image reconstruction, one popular class of methods exploits *synthesis dictionary learning* (DL).

1) Synthesis Dictionary Model: As briefly discussed in Section III, the synthesis model suggests that a signal can be approximated by a sparse linear combination of atoms or columns of a dictionary, i.e., the signal lives approximately in a subspace spanned by a few dictionary atoms. Because different signals may be approximated with different subsets of dictionary columns, the model is viewed as a union of subspaces model [106].

In imaging, the synthesis model is often applied to image patches (see Fig. 2) or image blocks $\boldsymbol{P}_j\boldsymbol{x}$ as $\boldsymbol{P}_j\boldsymbol{x} \approx \boldsymbol{D}\mathbf{z}_j$, with $\boldsymbol{P}_j$ denoting the operator that extracts a vectorized patch (with n pixels) of $\boldsymbol{x}$, $\boldsymbol{D} \in \mathbb{C}^{n\times K}$ denoting a synthesis dictionary (in general complex-valued), and $\mathbf{z}_j \in \mathbb{C}^K$ being the sparse representation or code for the patch $\boldsymbol{P}_j\boldsymbol{x}$ with many zeros. Although dictionaries based on the discrete cosine transform (DCT), etc., can be used to model image patches, much better representations can be obtained by adapting the dictionaries to data. The learning of synthesis dictionaries has been explored in many works [23], [107], [108] and shown to be promising in inverse problem settings [27], [109], [110].

2) Dictionary Learning for MRI: A DL-based method for MRI (DL-MRI) was proposed in [27], where the image and the dictionary for its patches are simultaneously estimated from limited measurements. The approach also known as blind CS (BCS) [30] does not require the training data and learns a dictionary that is highly adaptive to the underlying image content. However, the optimization problem is highly nonconvex and is formulated as follows:

$$\min_{\boldsymbol{x},\boldsymbol{D},\boldsymbol{Z}} \frac{1}{2}\|\boldsymbol{A}\boldsymbol{x}-\boldsymbol{y}\|_2^2 + \beta \sum_{j=1}^{N} \|\boldsymbol{P}_j\boldsymbol{x} - \boldsymbol{D}\mathbf{z}_j\|_2^2$$
$$\text{s.t. } \|\mathbf{z}_j\|_0 \le s, \quad \|\mathbf{d}_i\|_2 = 1, \ \forall i, j. \tag{14}$$

This corresponds to using a DL regularizer (weighted by $\beta > 0$) of the following form:

$$R(\boldsymbol{x}) = \min_{\boldsymbol{D},\boldsymbol{Z}} \sum_{j=1}^{N} \|\boldsymbol{P}_j \boldsymbol{x} - \boldsymbol{D}\mathbf{z}_j\|_2^2 \quad \text{s.t. } \|\mathbf{z}_j\|_0 \leq s, \quad \|\mathbf{d}_i\|_2 = 1, \ \forall i, j \tag{15}$$

where $\boldsymbol{Z}$ is a matrix whose columns are the sparse codes $\mathbf{z}_j$ that each have at most s nonzeros, and the ℓ_0 "norm" counts the total number of nonzeros in a vector or matrix. The columns $\mathbf{d}_i$ of $\boldsymbol{D}$ are constrained to have unit norm as, otherwise, $\mathbf{d}_i$ can be scaled arbitrarily along with corresponding inverse scaling of the ith row of $\boldsymbol{Z}$, and the objective is invariant to this scaling ambiguity.

Problem (14) was optimized in [27] by alternating between solving for the image $\boldsymbol{x}$ (image update step) and optimizing the dictionary and sparse coefficients (DL step). In specific cases such as in single-coil Cartesian MRI, the image update step is solved in closed-form using FFTs. However, the DL step involves a nonconvex and NP-hard optimization problem [111]. Various DL algorithms exist for this problem and its variants [23], [108], [112] that often alternate between updating the sparse coefficients (sparse coding) and the dictionary. The DL-MRI method for (14) used the K-SVD DL algorithm [23] and showed significant image quality improvements over previous CS MRI methods that used nonadaptive wavelets and TV [15]. However, it is slow due to expensive and repeated sparse coding steps and lacked convergence guarantees. In practice, variable rather than common sparsity levels across patches can be allowed in DL-MRI by using an error threshold-based stopping criterion when sparse coding with orthogonal matching pursuit (OMP) [113].

3) Other Applications and Variations: Later works applied DL to dynamic MRI [28], [114], [115], parallel MRI [116], and PET reconstruction [117]. An alternative Bayesian nonparametric DL approach was used for MRI reconstruction in [118]. DL was studied for CT image reconstruction in [25], which compared the BCS approach to prelearning the dictionary from a data set and fixing it during reconstruction. The former was found to be more promising when sufficient views (in sparse-view CT) were measured, whereas with very few views (or with very little measured information), prelearning performed better. Tensor-structured (patch-based) DL has also been exploited recently for dynamic CT [119] and spectral CT [120] reconstructions.

4) Recent Efficient Dictionary-Learning-Based Methods: Recent work proposed efficient DL-based reconstruction algorithms, dubbed sum of outer products dictionary learning (SOUP-DIL) image reconstruction algorithms [121] that used the following regularizer:

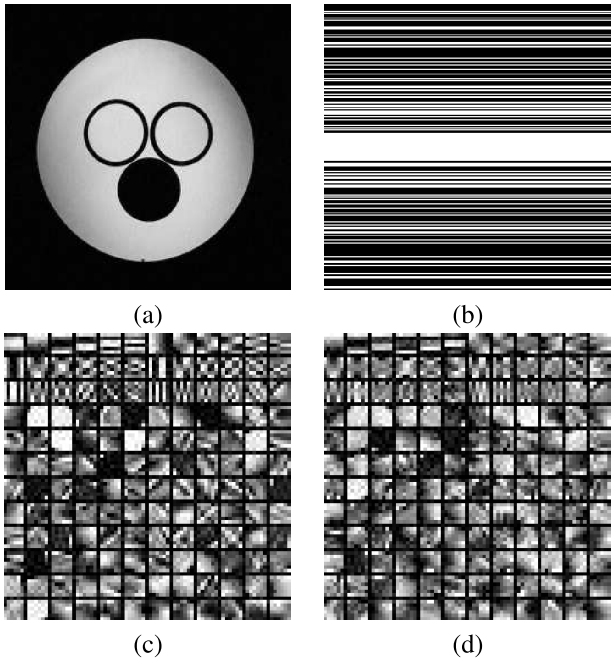

Fig. 3. *DL for MRI (images from [121]). (a) SOUP-DILLO MRI [121] reconstruction (uses SOUP-DIL with ℓ_0 penalty) of the water phantom [102]. (b) Sampling mask in k-space with 2.5× undersampling. (c) Real and (d) imaginary parts of the dictionary learned during reconstruction, with atoms shown as 6 × 6 patches.*

$$\min_{\boldsymbol{D},\boldsymbol{Z}} \sum_{j=1}^{N} \left\{ \|\boldsymbol{P}_j \boldsymbol{x} - \boldsymbol{D}\mathbf{z}_j\|_2^2 + \lambda^2 \|\mathbf{z}_j\|_0 \right\} \ \text{s.t. } \|\mathbf{d}_i\|_2 = 1, \ \forall i. \tag{16}$$

Here, the *aggregate* sparsity penalty $\sum_{j=1}^{N} \|\mathbf{z}_j\|_0$ with weight λ^2 automatically enables variable sparsity levels across patches. The DL step of the SOUP-DIL reconstruction algorithm efficiently optimized (16) using an exact block coordinate descent (BCD) scheme by decomposing $\boldsymbol{DZ}$ as a sum of outer products (SOUP) of dictionary columns and rows of $\boldsymbol{Z}$, and solving for $\mathbf{d}_i$ and then the ith row of $\boldsymbol{Z}$ (by thresholding) in closed-form, and cycling over all such pairs ($1 \leq i \leq K$).

Although the earlier DL-MRI used inexact (greedy) and expensive sparse code updates and lacked convergence analysis, the SOUP-DIL scheme used efficient, exact updates and was proven to converge to the critical points (generalized stationary points) of the underlying problems and improved image quality over several schemes [121]. Fig. 3 shows an example reconstruction with this BCS method along with the learned dictionaries. Another recent work [122] extended the L + S model for dynamic image reconstruction in (7) to a low rank and adaptive sparse signal model that incorporated a DL regularizer similar to (16) for the $\boldsymbol{x}_S$-component.

5) Alternative Convolutional Dictionary Model: One can replace the patch-based dictionary model with a convolutional model as $\boldsymbol{x} \approx \sum_{i=1}^{K} \mathbf{d}_i \otimes \mathbf{c}_i$ that directly represents the image as a sum of (possibly circular) convolutions of

dictionary filters $\mathbf{d}_i$ and sparse coefficient maps $\mathbf{c}_i$ [123], [124]. The convolutional synthesis dictionary model is distinct from the patch-based model. However, its main drawback is the inability to represent very low-frequency content in images, necessitating preprocessing of images to remove very low-frequency content prior to convolutional DL. The utility of convolutional synthesis DL for biomedical image reconstruction is an open and interesting area for future research (see [125] for a denoising formulation that could be extended to inverse problems).

C. Sparsifying Transform Learning-Based Methods

Several recent works have studied the learning of the efficient *sparsifying transform model* for biomedical image reconstruction [26], [29], [126]. This section reviews these advances (see [127] for an MRI focused review).

1) Transform Model: The sparsifying transform model is a generalization [24] of the analysis dictionary model. The latter assumes that applying an operator $\boldsymbol{W}$ to a signal $\mathbf{f}$ produces several zeros in the output, i.e., the signal lies in the null space of a subset of rows of the operator. The sparsifying transform model allows for a sparse approximation as $\boldsymbol{W}\mathbf{f} = \mathbf{z} + e$, where $\mathbf{z}$ has several zeros and e is a transform-domain modeling error. Natural images are well known to be approximately sparse in transform domains such as the DCT and wavelets, a property that has been exploited for image compression [128], denoising, and inverse problems. A key advantage of the sparsifying transform model compared to the synthesis dictionary model is that the transform-domain sparse approximation can be computed exactly and cheaply by thresholding $\boldsymbol{W}\mathbf{f}$ [24].

2) Early Efficient Transform Learning-Based Methods: Recent works [29], [126] proposed transform learning-based image reconstruction methods that involved computationally cheap, closed-form updates in the iterative algorithms. The following square transform learning [24] regularizer was used for reconstruction in [126]:

$$R(\boldsymbol{x}) = \min_{\boldsymbol{W},\boldsymbol{Z}} \sum_{j=1}^{N} \|\boldsymbol{W}\boldsymbol{P}_j\boldsymbol{x} - \mathbf{z}_j\|_2^2 + \gamma Q(\boldsymbol{W}) \ \text{ s.t. } \ \|\boldsymbol{Z}\|_0 \leq s \tag{17}$$

where $\boldsymbol{W} \in \mathbb{C}^{n\times n}$ is a square matrix and the transform learning regularizer $Q(\boldsymbol{W}) = -\log|\det \boldsymbol{W}| + 0.5\|\boldsymbol{W}\|_F^2$ with weight $\gamma > 0$ prevents trivial solutions in learning such as the zero matrix or matrices with repeated rows. Moreover, it also helps to control the condition number of the transform [24]. The term $\sum_{j=1}^{N}\|\boldsymbol{W}\boldsymbol{P}_j\boldsymbol{x} - \mathbf{z}_j\|_2^2$ denotes the transform-domain modeling error or sparsification error, which is minimized to learn a good sparsifying transform. The constraint in (17) on the ℓ_0 "norm" of the matrix $\boldsymbol{Z}$ controls the net or aggregate sparsity of all patches' sparse coefficients.

The image reconstruction problem with regularizer (17) was solved in [126] using a highly efficient BCD approach that alternates between minimizing with respect to $\boldsymbol{Z}$ (transform sparse coding step), $\boldsymbol{W}$ (transform update step), and $\boldsymbol{x}$ (image update step). Importantly, the transform sparse coding step has a closed-form solution, where the matrix $\boldsymbol{B}$, whose columns are $\boldsymbol{W}\boldsymbol{P}_j\boldsymbol{x}$, is thresholded to its s largest magnitude elements, with other entries set to zero. When the sparsity constraint is replaced with alternative sparsity promoting functions such as the ℓ_0 sparsity penalty or ℓ_1 penalty, the sparse coding solution is obtained in closed form by hard or soft thresholding. The transform update step has a simple solution involving the singular value decomposition (SVD) of a small matrix [126], and the image update step involves a least-squares problem (e.g., in the case of single-coil Cartesian MRI, it is solved in closed form using FFTs [126]). This efficient BCD scheme was proven to converge, in general, to the critical points of the nonconvex reconstruction problem [126].

In practice, the sparsity controlling parameter can be varied over algorithm iterations (a continuation strategy), allowing for faster artifact removal initially and then reduced bias over the iterations [29]. The scheme in [126] was shown to be much faster than the previous DL-MRI scheme. Tanc and Eksioglu [129] further combined transform learning with global sparsity regularization in known transform domains for CS MRI.

Square transform learning has also been applied to CT reconstruction [130]. Another recent work used square transform learning for low-dose CT image reconstruction [131] with a shifted-Poisson likelihood penalty for the data-fidelity term in the cost (instead of the conventional weighted least-squares penalty), but prelearned the transform from a data set and fixed it during reconstruction to save computation.

Other works have explored alternative formulations for transform learning (e.g., overcomplete or tall [132] transforms) that could be potentially used for image reconstruction.

3) Learning Rich Unions of Transforms for Reconstruction: Since images typically contain a diversity of textures, features, and edge information, recent works [26], [29], [133] learned a union of transforms (a rich model) for image reconstruction. In this setting, a collection of K transforms is learned, and the image patches are grouped or *clustered* into K classes, with each class of (*similar*) patches best matched to and using a particular transform. The union of transforms learning (UNITE) image reconstruction formulation in [29] uses the following regularizer:

$$R(\boldsymbol{x}) = \min_{\{\boldsymbol{W}_k, C_k, \mathbf{z}_j\}} \sum_{k=1}^{K}\sum_{j\in C_k} \left\{\|\boldsymbol{W}_k\boldsymbol{P}_j\boldsymbol{x} - \mathbf{z}_j\|_2^2 + \lambda^2\|\mathbf{z}_j\|_0\right\}$$
$$\text{s.t. } \boldsymbol{W}_k^H\boldsymbol{W}_k = \mathbf{I} \quad \forall k,\ \{C_k\} \in G. \tag{18}$$

Here, C_k is a set containing the indices of all patches matched to the transform $\boldsymbol{W}_k$, and G denotes the set of

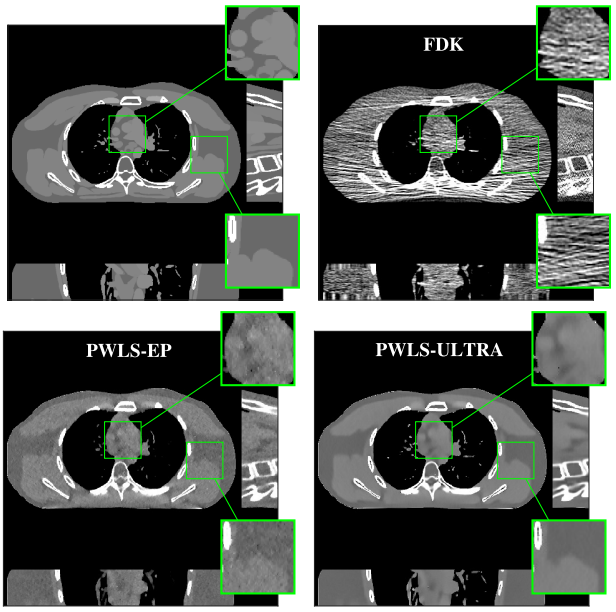

Fig. 4. *Cone-beam CT reconstructions (images from [26]) of the extended cardiac-torso (XCAT) phantom [134] using the FDK, PWLS-EP [135] (with EP regularizer), and PWLS-ULTRA [26] ($K = 15$) methods at dose $I_0 = 5 \times 10^3$ incident photons per ray, shown along with the ground truth (top left). The central axial, sagittal, and coronal planes of the 3-D reconstruction are shown. The learning-based PWLS-ULTRA removes noise and preserve edges much better than the other schemes.*

all partitions of $[1 : N]$ into K disjoint subsets, where N is the total number of overlapping patches. Note that when indexed variables are enclosed in braces [in (18) and later equations], we mean the set of all variables over the range of the indices.

The UNITE reconstruction formulation jointly learns a collection of transforms, clusters and sparse codes patches, and reconstructs the image $\boldsymbol{x}$ from measurements. An efficient BCD algorithm with convergence guarantees was proposed for optimizing the problem in [29]. The K transforms in (18) are unitary, which simplifies the BCD updates. For MRI, UNITE-MRI achieved improved image quality over the square transform learning-based scheme when reconstructing from undersampled k-space measurements [29].

Recent works applied learned unions of transforms to other applications. For example, the union of transforms model was prelearned (from a data set) and used in a clustering-based low-dose 3-D CT reconstruction scheme [26]. Fig. 4 shows an example of high quality reconstructions obtained with this scheme. Although the work used a PWLS-type reconstruction cost, a more recent method [131] replaced the weighted least-squares data-fidelity term with the shifted-Poisson likelihood penalty, which further improved the image quality and reduced bias in the reconstruction in ultralow-dose settings. Other recent works combined learned union of transforms models with material image models and applied it to image-domain material decomposition in dual-energy CT with high quality results [136], [137].

4) Learning-Structured Transform Models: It is often useful to incorporate various structures and invariances in learning to better model the natural data and to prevent learning spurious features in the presence of noise and corruptions. Flipping and rotation invariant sparsifying transform learning was recently proposed and applied to image reconstruction in [138]. The regularization is similar to (18), but using $\boldsymbol{W}_k = \boldsymbol{W}\boldsymbol{\Phi}_k$ with a common *parent transform* $\boldsymbol{W}$ and $\{\boldsymbol{\Phi}_k\}$ denoting a set of known flipping and rotation operators that apply to each (row) atom of $\boldsymbol{W}$ and approximate flips and rotations by permutations (similar to [101], but which used fixed 1-D Haar wavelets as the parent). This enables learning a much more structured but flexible (depending on the number of operators $\boldsymbol{\Phi}_k$) model than in (18), with clustering done more based on similar directional properties. Images with more directional features are better modeled by such learned transforms [138].

5) Learning Complementary Models—Low-Rank and Transform Sparsity: A recent work [139] proposed an approach called Sparsifying TRansfOrm Learning and Low-Rank (STROLLR) that combines two complementary regularizers: one exploiting (nonlocal) self-similarity between regions and another exploiting transform learning that is based on local patch sparsity. Nonlocal similarity and block matching models are well known to have excellent performance in image processing tasks such as image denoising (with block-matching and 3-D filtering or BM3D [140]). The STROLLR regularizer has the form $R(\boldsymbol{x}) = R_1(\boldsymbol{x}) + R_2(\boldsymbol{x})$, where the low-rank regularizer is as follows:

$$R_1(\boldsymbol{x}) = \min_{\{\mathbf{U}_j\}} \sum_{j=1}^{N} \left\{ \|\mathcal{M}_j(\boldsymbol{x}) - \mathbf{U}_j\|_F^2 + \eta^2 \operatorname{rank}(\mathbf{U}_j) \right\} \quad (19)$$

and the transform learning regularizer is

$$R_2(\boldsymbol{x}) = \min_{\boldsymbol{W}, \{\mathbf{z}_j\}} \sum_{j=1}^{N} \left\{ \|\boldsymbol{W}\mathbf{H}_j\boldsymbol{x} - \mathbf{z}_j\|_2^2 + \lambda^2 \|\mathbf{z}_j\|_0 \right\} \quad \text{s.t. } \boldsymbol{W}^H\boldsymbol{W} = \mathbf{I}. \quad (20)$$

Here, the operator $\mathcal{M}_j$ is a block matching operator that extracts the jth patch $\boldsymbol{P}_j\boldsymbol{x}$ and the $L-1$ patches most similar to it and forms a matrix, whose columns are the jth patch and its matched siblings, ordered by the degree of match. This matrix is approximated by a low-rank matrix $\mathbf{U}_j$ in (19), with $\eta > 0$. The vector $\mathbf{H}_j\boldsymbol{x}$ is a vectorization of the submatrix that is the first P columns of $\mathcal{M}_j(\boldsymbol{x})$. Thus, the regularizer in (20) learns a higher dimensional transform (e.g., 3-D transform for 2-D patches) and jointly sparsifies nonlocal but similar patches.

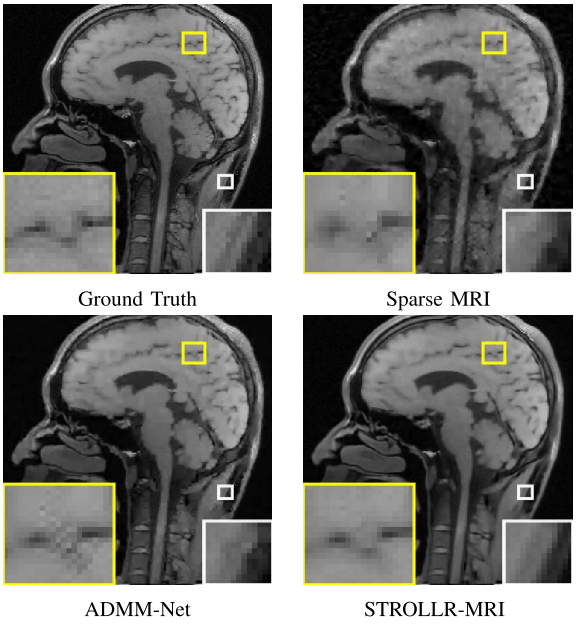

Fig. 5. *MRI reconstructions (images from [127]) with pseudo-radial sampling and 5× undersampling using Sparse MRI [15] (peak signal-to-noise ratio or PSNR = 27.92 dB), ADMM-Net [141] (PSNR = 30.67 dB), and STROLLR-MRI [139] (PSNR = 31.98 dB), along with the original image from [141]. STROLLR-MRI clearly outperforms the nonadaptive Sparse MRI, while ADMM-Net also produces undesirable artifacts.*

STROLLR-MRI [139] was shown to achieve better CS MRI image quality over several methods including the supervised (deep) learning-based alternating direction method of multipliers (ADMM)-Net [141]. Fig. 5 shows example MRI reconstructions and comparisons. Similar to UNITE-MRI, there is an underlying grouping of patches, but STROLLR-MRI exploits block-matching and sparsity to implicitly perform grouping.

D. Online Learning for Reconstruction

Recent works have proposed online learning of sophisticated models for reconstruction, particularly of dynamic data from time-series measurements [32], [142], [143]. In this setting, the reconstructions are produced in a time-sequential manner from the incoming measurement sequence, with the models also adapted simultaneously and sequentially over time to track the underlying object's dynamics and aid reconstruction. Such methods allow greater adaptivity to temporal dynamics and can enable dynamic reconstruction with less latency, memory use, and computation than conventional methods. Potential applications include real-time medical imaging, interventional imaging, etc., or they could be used even for more efficient and (spatially and temporally) adaptive offline reconstruction of large-scale (*big*) data.

A recent work efficiently adapted low-rank tensor models in an online manner for dynamic MRI [143]. Online learning for dynamic image reconstruction was shown to be promising in [32] and [143], which adapted synthesis dictionaries to spatio-temporal patches. In this setup [32], measurements corresponding to a group (called minibatch) of frames are processed at a time using a sliding window strategy. The objective function for reconstruction is a weighted time average of instantaneous cost functions, each corresponding to a group of processed frames. An exponential weighting (forgetting) factor for the instantaneous cost functions controls the past memory in the objective. The instantaneous cost functions include both a data-fidelity and a regularizer (corresponding to patches in the group of frames) term. The objective function thus changes over time and is optimized at each time point with respect to the most recent minibatch of frames and corresponding sparse coefficients (with older frames and coefficients fixed), but the dictionary is itself adapted therein to all the data. Each frame can be reconstructed from multiple overlapping temporal windows and a weighted average of those used as the final estimate.

The online learning algorithms in [32] achieved computational efficiency by using warm start initializations (that improve over time) for variables and frames based on estimates in previous windows, and thus running only a few iterations of optimization for each new window. They stored past information in small (cumulatively updated) matrices for the dictionary update (low memory usage). The methods were significantly more efficient and more effective than batch learning-based techniques for dynamic MRI that iteratively learn and reconstruct from *all* k-t space measurements. Given the potential of online learning methods to transform dynamic and large-scale imaging, we expect to see growing interest and research in this domain.

E. Connections Between Transform Learning Approaches and Convolutional Network Models

The sparsifying transform models in Section V-C have close connections with convolutional filterbanks. This section and the next review some of these connections and implications for reconstruction.

1) Connections to Filterbanks: Transform learning and its application to regularly spaced image patches [144] can be equivalently performed using convolutional operations. For example, applying an atom of the transform to all the overlapping patches of an (2-D) image via inner products is equivalent to convolving the image with a transform filter that is the (2-D) flipped version of the atom. Thus, sparse coding in the transform model can be viewed as convolving the image with a set of transform filters (obtained from the transform atoms) and thresholding the resulting filter coefficient maps, and transform learning can be viewed as equivalently learning convolutional sparsifying filters [145], [146]. When using only a regularly spaced subset of patches, the above interpretation of transform sparse coding modifies to convolving the image with the transform filters, downsampling the results, and then thresholding [144]. Transform models based on clustering [29] add nontrivial complexities to this process.

Applying the matrix $\boldsymbol{W}^H$ to the sparse codes of all overlapping patches and spatially aggregating the results an operation used in iterative transform-based reconstruction algorithms [29], is equivalent to filtering the thresholded filter coefficient maps with corresponding matched filters (complex conjugate of transform atoms) and summing the results over the channels. These equivalences between patch-based and convolutional operations for the transform model contrast with the case for the synthesis dictionary model in Section V-B, where the patch-based and convolutional versions of the model are not equivalent in general. When a disparate set of (e.g., randomly chosen) image patches or operations such as block matching [139], etc., are used with the transform model, the underlying operations do not correspond to convolutions (thus, the transform learning frameworks can be viewed as more general). Typically, the convolutional implementation of transforms is more computationally efficient than the patch-based version for large filter sizes [144].

Recent works have exploited the filterbank interpretation of the transform model [144], [147], [148]. For example, in [149], the authors learned filterbanks for MRI reconstruction. In [144], they studied alternative properties and regularizers for transform learning.

2) Multilayer Transform Learning: A recent work [148] proposed learning multilayer extensions of the transform model [dubbed deep residual transforms (DeepResT)] that mimic CNNs by incorporating components such as filtering, nonlinearities, pooling, and stacking; however, the learning was done using unsupervised *model-based transform learning-type cost functions*.

In the conventional transform model, the image is passed through a set of transform filters and thresholded (the nonlinearity) to generate the sparse coefficient maps. In the DeepResT model, the residual (difference) between the filter outputs and their sparse versions is computed, and these residual maps for different filters are stacked together to form a residual volume that is jointly sparsified in the next layer. To prevent dimensionality explosion, each filtering of the residual volume in the second and subsequent layers produces a 2-D output (for a 2-D initial image). The multilayer model thus consists of successive joint sparsification of residual maps several times (cf. [149, Fig. 1] and [128, Fig. 9]). The filters and sparse maps in all layers of the (*encoder*) network are jointly and efficiently learned in [148] from images to provide the smallest sparsification residuals in the final (output) layer, a transform learning-type cost. The learned model and multilayer sparse coefficient maps can then be backpropagated in a linear fashion (*decoder*) to generate image approximations. The DeepResT model also downsampled (pooled) the residual maps (along the filter channel dimension) in each encoder layer before further filtering them, providing robustness to noise and data corruptions. The learned models [148] were shown to provide promising performance for denoising images when learning directly from the noisy data, and moreover learning stacked multilayer encoder–decoder modules was shown to improve the performance, especially at high noise levels. Application of such deep transform models to medical image reconstruction [150] is an ongoing area of potent research.

F. Physics-Driven Deep Training of Transform-Based Reconstruction Models

There has been growing recent interest in supervised learning approaches for image reconstruction [33]. These methods learn the parameters of reconstruction algorithms from training data sets (typically consisting of pairs of ground-truth images and initial reconstructions from measurements) to minimize the error in reconstructing the training images from their typically limited or corrupted measurements. For example, the reconstruction model can be a deep CNN (typically consisting of encoder and decoder parts) that can be trained (as a denoiser) to produce a reconstruction from an initial corrupted version [34]. Section VI discusses such approaches in more detail. These methods can often require large training sets to learn billions of parameters (e.g., filters, etc.). Moreover, learned CNNs (*deep learning*) may not typically or rigorously incorporate the imaging measurement model or the information about the physics of the imaging process, which are a key part of solving inverse problems. Hence, there has been recent interest in learning the parameters of iterative algorithms that solve regularized inverse problems [35], [141] (cf. Section VI for more such methods). These methods can also typically have fewer free parameters to train.

Recent works have interpreted early transform-based BCS algorithms as deep physics-driven convolutional networks learned on-the-fly, i.e., in a blind manner, from measurements [35], [147]. For example, the image update step in the square transform BCS (that learns a unitary transform) algorithm in [29] involves a least-squares-type optimization with the following normal equation:

$$\boldsymbol{G}\boldsymbol{x}^k = \nu \boldsymbol{A}^H \boldsymbol{y} + \sum_{j=1}^{N} \boldsymbol{P}_j^T \boldsymbol{D}^k \boldsymbol{H}_\lambda(\boldsymbol{W}^k \boldsymbol{P}_j \boldsymbol{x}^{k-1}) \qquad (21)$$

where $\nu = 1/\beta$ (for β in Section V-B or V-C) and k denotes the iteration number in the BCD reconstruction algorithm. Matrix $\boldsymbol{D}^k \triangleq (\boldsymbol{W}^k)^H$ is a (matched) synthesis operator, and $\boldsymbol{G} \triangleq \sum_{j=1}^{N} \boldsymbol{P}_j^T \boldsymbol{P}_j + \nu \boldsymbol{A}^H \boldsymbol{A}$ is a fixed matrix. The hard-thresholding in (21) corresponds to the solution of the sparse coding step of the BCD algorithm [29].

Fig. 6 shows an unrolling of K iterations (layers) of (21), with fresh filters in each iteration. Each layer has a system model block that solves (21) [e.g., with FFTs or the conjugate gradient (CG) method], whose inputs are the two terms on the right-hand side of (21): the first term is a *fixed bias* term; and the second term (denotes a decorruption step) is computed via convolutions by first applying the transform filters (denoted by h_l^k, $1 \leq l \leq L$

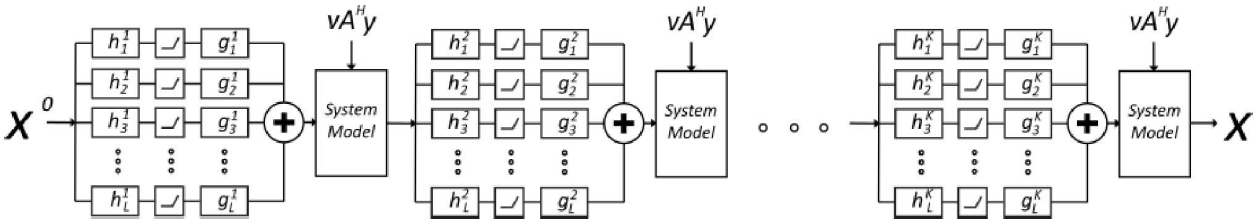

Fig. 6. *Reconstruction model (see [147]) derived from the image update step of the square transform learning-based image reconstruction algorithm in [29]. The model here has K layers corresponding to K iterations. Each layer first has a decorruption step that computes the second term in (21) using filtering and thresholding operations, assuming a transform model with L filters. This is followed by a system model block that adds the fixed bias term $\nu A^H y$ to the output of the decorruption step and performs a least-squares-type image update (e.g., using CG) to enforce the imaging forward model.*

in Fig. 6) followed by thresholding (the nonlinearity) and then matched synthesis filters (denoted by g_l^k, $1 \leq l \leq L$), and summing the outputs over the filters. This is clear from writing the second term in (21) as $\sum_{l=1}^{L} \sum_{j=1}^{N} \boldsymbol{P}_j^T \mathbf{d}_l^k \boldsymbol{H}_\lambda(\mathbf{r}_l^{k^T} \boldsymbol{P}_j \boldsymbol{x}^{k-1})$, with $\mathbf{d}_l$ and $\mathbf{r}_l$ denoting the lth columns of $\boldsymbol{D}$ and $\mathbf{R} = \boldsymbol{W}^T$, respectively. Each of the L terms forming the outer summation here corresponds to the output of an arm (of transform filtering, thresholding, and synthesis filtering) in the decorruption module shown in Fig. 6. Since the BCS scheme in [29] does not use training data, but rather learns the transform filters as part of the iterative BCD algorithm (hence, the transform could change from iteration to iteration or layer to layer), it can be interpreted as learning the model in Fig. 6 in an on-the-fly sense from measurements.

Recent works [35], [147] learned the filters in this multilayer model (BCD Net [151]) with soft-thresholding (ℓ_1-norm-based) nonlinearities and trainable thresholds using a greedy scheme to minimize the error in reconstructing a training set from limited measurements. These and similar approaches (including the transform-based ADMM-Net [141]) involving unrolling of typical image reconstruction algorithms are *physics-driven deep training* methods due to the systematic inclusion of the imaging forward model in the convolutional network. Once learned, the reconstruction model can be efficiently applied to test data using convolutions, thresholding, and least-squares-type updates. While the works in [35] and [148] did not enforce the corresponding synthesis and transform filters (in each arm of the decorruption module) to be matched in each layer, recent work [151] learned matched filters, improving image quality. The learning of such physics-driven networks is an active area of research, with interesting possibilities for new innovation in the convolutional models in the architecture motivated by more recent transform and DL-based (or other) reconstruction methods. In such methods, the thresholding operation is the key to exploiting sparsity.

VI. DEEP LEARNING METHODS

One of the most important recent developments in the field of image reconstruction is the introduction of deep learning approaches [38]. Motivated by the tremendous success of deep learning for image classification [152], [153], image segmentation [154], denoising [155], etc., many groups have recently successfully applied deep learning approaches to various image reconstruction problems such as in X-ray CT [156]–[162], MRI [33], [34], [161], [163]–[166], PET [167], [168], ultrasound [169], [170], and optics [171]–[173]. As of mid 2019, two commercial CT vendors received FDA approval for deep learning image reconstruction [174], [175].

The sharp increase in deep learning approaches for image reconstruction problems may be due to the "perfect storm" resulting from a combination of multiple attributes in perfect timing: availability of large public data, well-established GPU infrastructure in the image reconstruction community, easy-to-access deep learning toolboxes, industrial push, and open publications using arXiv.

For example, one important public data set that has significantly contributed to this wave is the 2016 American Association of Physicists in Medicine (AAPM) Low-Dose X-ray CT Grand Challenge data set [176]. The training data sets consist of normal-dose and quarter-dose abdominal CT data from ten patients. Another emerging important data set is the fast MRI data set by NYU Langone Health and Facebook [177]. The data set comprises raw k-space data from more than 1500 fully sampled knee MRIs and Digital Imaging and Communications in Medicine (DICOM) images from 10 000 clinical knee MRIs obtained at 3 or 1.5 T.

In addition, GPU methods have been extensively implemented to accelerate iterative methods in the field of image reconstruction. As a result, open deep learning toolboxes such as Tensorflow, pyTorch, MatConvNet, etc., based on GPU programming, are easily accessible to researchers in the field of image reconstruction. Moreover, the industry has been engaging in a big push in this development from the early phase, since deep learning-based image reconstruction methods are well suited to their business models. This is because the training can be done by the vendors with large databases and the users could enjoy high-quality reconstruction results at near real-time reconstruction speed.

Given the relatively long publication cycle for regular journals in the field of image reconstruction, most new developments are found in arXiv preprints, well before

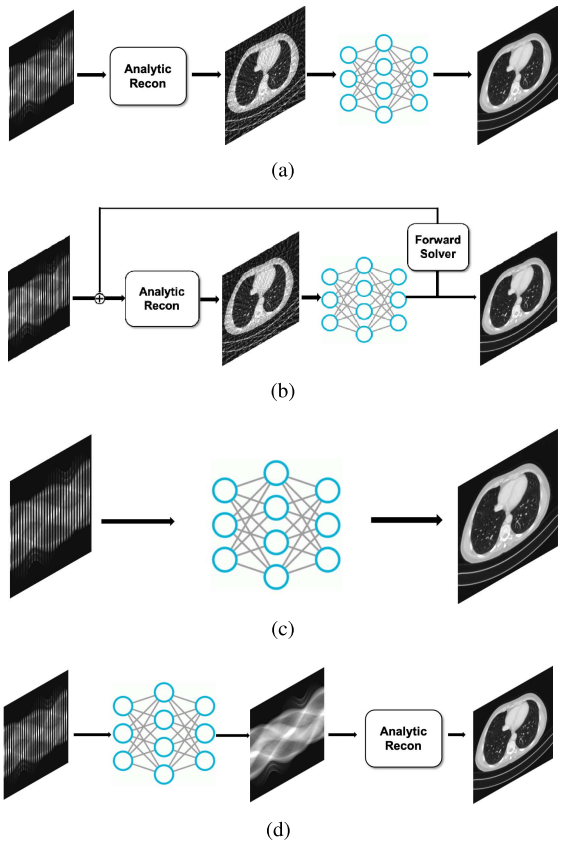

Fig. 7. *Various realizations of deep learning for image reconstruction. (a) Image-domain learning. (b) Hybrid-domain learning. (c) AUTOMAP. (d) Sensor-domain learning.*

they are formally accepted by the journals. This new trend of open publication facilitates the significant progress in this area within a short time period. The following section reviews the recent developments based on the peer-reviewed publications and arXiv preprints.

A. Categories of the Existing Approaches

This section starts with reviews of various network architectures and designs principles for image reconstruction problems. Fig. 7 illustrates typical architectures at a high level, which are commonly used in the literature. However, this field is rapidly growing, so a complete review of the design principles is beyond the scope of this article.

1) Image-Domain Learning: In image-domain approaches [156], [157], [160]–[162], [164], [165], [178], [179], artifact-corrupted images are first generated from the measurement data using some analytic methods (e.g., FBP, Fourier transform, etc.), from which NNs are trained to learn the artifacts [see Fig. 7(a)]. For example, the low-dose and sparse CT NNs [156], [157], [160]–[162], [179] belong to this class, where the noise corrupted images are first generated from the noisy or sparse view sinogram data using FBP, after which the artifacts are learned by comparing with the noiseless label images. In MR applications, the early U-Net architectures for CS MRI [165], [180] were also designed to remove the aliasing artifacts after obtaining the Fourier inversion image from the downsampled k-space data.

In particular, FBPConvNet [161] showed that image-domain networks can be derived by unrolling sparse recovery for one specific class of inverse problems: those where the normal operator associated with the forward model is a convolution. For this class of inverse problems, a CNN then emerges. This class of normal operators includes MRI, parallel-beam X-ray CT, and diffraction tomography (DT).

A current trend in image-domain learning is to use more sophisticated loss functions to overcome the observed smoothing artifacts. For example, Yang *et al.* [181] used the perceptual loss and the Wasserstein distance loss to improve resolution.

2) Hybrid-Domain Learning: In this class of approaches [33], [141], [158], [159], [163]–[165], [182]–[187], the data consistency term is imposed in the NN training and inference to improve the performance as shown in Fig. 7(b). The physics-driven deep training methods in Section V-F included the full data-fidelity-based image update in each layer.

The learned iterative soft-thresholding (LISTA) [188] is one of the earliest unfolding approaches that uses a time unfolded version of the ISTA algorithm [189]. Specifically, the weight matrices and sparsifying soft-thresholding operator are learned from the data (as also done in later works discussed in Section V-F).

The variational NN for CS MRI [164] derived an unrolled NN that uses a data consistency term for each layer. Specifically, the variational network is based on unfolding the following optimization problem:

$$\min_{\boldsymbol{x}} \frac{\lambda}{2}\|\boldsymbol{y}-\boldsymbol{A}\boldsymbol{x}\|^2 + \mathcal{R}(\boldsymbol{x}) \tag{22}$$

where the regularization term $\mathcal{R}(\boldsymbol{x})$ is represented as a sum of multichannel operations

$$\mathcal{R}(\boldsymbol{x}) = \sum_{i=1}^{M} \langle \Phi_i(\boldsymbol{K}_i\boldsymbol{x}), \boldsymbol{1}\rangle \tag{23}$$

where $\boldsymbol{K}_i$ denotes the ith linear operator represented by the ith channel convolution (like the transform filters in Section V), and Φ_i denotes the associated activation function. In a variational network [164], the convolution-based linear operator $\boldsymbol{K}_i$, the gradient of the activation function Φ_i', and the regularization parameter λ^k are learned for each unfolded Landweber iteration steps. Related approaches have been taken in dynamic cardiac MRI [33].

In ADMM-Net [141], the unrolled steps of the ADMM-based reconstruction algorithm are mapped to each layer of a deep NN. Recently, the primal–dual algorithm was extended to obtain a CNN-based learned primal–dual approach [159]. In this approach, two

NNs are learned for the primal step and dual step. The projected gradient method has also been extended to a NN approach [182].

Another class of popular hybrid-domain approaches is based on the CNN penalty and plug-and-play model. Specifically, in CNN penalty approaches [163], a NN is used as a prior model within an MBIR framework. Rather than using a CNN penalty explicitly, in the plug-and-play approach [186], [187], the denoising step of an iteration like ADMM is replaced with a NN denoiser. Similarly, the deep image prior approach [190] formulates the image reconstruction problem as

$$\min_{\boldsymbol{\theta}} \frac{\lambda}{2}\|\boldsymbol{y} - \boldsymbol{A}\boldsymbol{z}(\boldsymbol{\theta})\|^2 + \mathcal{R}(\boldsymbol{z}(\boldsymbol{\theta})) \tag{24}$$

$$\text{s.t. } \boldsymbol{z}(\boldsymbol{\theta}) = \boldsymbol{G}_{\boldsymbol{\theta}}(\mathbf{v}) \tag{25}$$

where $\boldsymbol{G}_{\boldsymbol{\theta}}$ is a deep NN parameterized by $\boldsymbol{\theta}$. The dimension of the parameter $\boldsymbol{\theta}$ is usually determined by the NN architecture. In the original deep prior model [190], the input ($\mathbf{v}$) for the NN was a noise vector. Instead, in its recent application to PET image reconstruction [168], simultaneously acquired MRI data were used as the input to the NN for PET reconstruction using deep image prior.

3) Automap: The Automated Transform by Manifold Approximation (AUTOMAP) [166] [see Fig. 7(c)] approach learns a direct mapping from the measurement domain to image domain using a NN. This approach requires a fully connected layer followed by convolution layers, leading to high memory requirements for storing that fully connected layer, currently limiting AUTOMAP applications to small size reconstruction problems in MRI.

4) Sensor-Domain Learning: Sensor-domain learning approaches try to learn the sensor-domain interpolation and denoising using an NN as shown in Fig. 7(d). For low-dose CT, the work in [191] designed an NN in the projection domain, yet the NN is trained in an end-to-end manner from the sinogram to the image domain. Accordingly, the final output of the NN is a data-driven ramp filter designed by minimizing the image-domain loss. Metal artifact correction in CT is another opportunity for sensor-domain learning [192], [193]. For low-dose CT and sparse view CT, direct sonogram-domain processing using deep NN was also proposed [194], [195]. In k-space deep learning for accelerated MRI [196]–[198], NNs were designed to learn k-space interpolation kernels in an end-to-end manner from k-space to the image domain using an image-domain loss.

5) Some Variations: In [156], the NNs were designed to learn the relationship between contourlet transform coefficients of the low-dose input and high-dose label data. Later, this problem is formally extended to wavelet-domain residual network (WavResNet) to improve the performance [158] (see Fig. 8). Here, the choice of appropriate transform domain facilitating efficient learning is important and usually based on domain expertise. For example, in a recent deep NN architecture for interior tomography problems, [199] observed that the NN is more robust with respect to different region of interest (ROI) sizes, detector pitch, short scan, and sparse view artifacts, if the NN is designed in the differentiated backprojection (DBP) domain. The DBP is well-known in the CT community for its robustness to short scan artifact, interior tomography, etc., which clearly shows the importance of domain expertise in designing NNs.

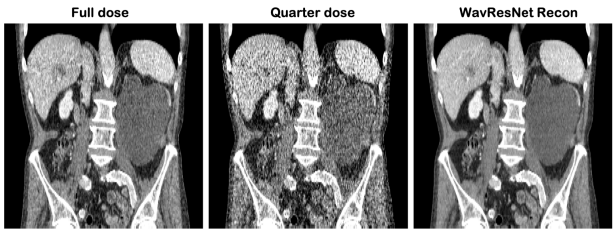

Fig. 8. *Left to right: full-dose FBP reconstruction, quarter-dose FBP reconstruction, and WavResNet denoising results [158] applied to 25% dose FBP images. The detailed textures, vessel structure, and cancer lesions are clearly seen in the denoised results.*

B. Semisupervised and Unsupervised Learning

Most deep learning approaches for image reconstruction have been based on the supervised learning framework. For example, in the low-dose CT reconstruction problems, the NN is trained to learn the mapping between the noisy image and the noiseless (or high dose) label images. Similar approaches are taken in accelerated MRI, where the relationship between highly accelerated and artifact corrupted input and the fully sampled label data are learned using training data.

Unfortunately, in many imaging scenarios, the noiseless label images are difficult to obtain or even impossible to acquire. For example, in low-dose CT problems, an institutional reviewer board (IRB) rarely approves experiments that would require two exposures at low- and high-dose levels due to the potential risks to patients. This is why in the AAPM X-ray CT Low-Dose Grand Challenge, the matched low-dose images were generated by adding synthetic noise to the full-dose sinogram data. Even in accelerated MRI, high-resolution fully sampled k-space data are very difficult to acquire due to the long scan time and impossible to collect for dynamic MRI data sets that are all inherently undersampled. Therefore, NN training without reference or with small reference pairs is very important in the field of image reconstruction.

One of the earliest works in this regard was a low-dose CT denoising network using generative adversarial network (GAN) loss [160]. Instead of using matched high-dose data, the authors employ the GAN loss to match the probability distribution. One of the limitations of this work is that the network is very sensitive and, without

careful training, spurious artifacts are often generated due to the generative nature of GAN. To address this problem, a cycleGAN architecture employed cyclic loss and identity loss for multiphase cardiac CT problems [200]. Thanks to the two-loss functions, the authors demonstrated that no spurious artifact appeared in their results even without reference data.

In addition, several of the learning approaches in Section V also do not require reference data (or can use limited reference data) to provide high-quality reconstructions.

C. Interpretation of Deep Models

One of the major hurdles of the deep learning approaches for image reconstruction is the black-box nature of NNs. This is especially problematic for medical imaging applications, since many clinicians are concerned about whether the performance improvement is real or cosmetic.

The modern reconstruction techniques, such as CS, can be considered as *representation learning* methods that aim at finding the optimal parsimonious representation under the data-fidelity term. Unfortunately, the classical approaches usually require computationally expensive optimization methods to find the optimal representation. One of the important take-home messages of this section is to show that deep learning approaches are indeed another form of representation learning approaches, which have advantages compared to the classical approaches. In the following, we start to revisit the classical approaches in this aspect.

1) Image Reconstruction via Representation Learning: One can formulate image reconstruction problems as

$$\min_{\boldsymbol{x}\in\boldsymbol{\mathcal{X}}} \quad \|\boldsymbol{y}-\boldsymbol{A}\boldsymbol{x}\|^2 \tag{26}$$

where $\boldsymbol{\mathcal{X}}$ denotes the low-dimensional manifold where the unknown $\boldsymbol{x}$ lives. For example, $\boldsymbol{\mathcal{X}}$ can be represented by

$$\boldsymbol{\mathcal{X}} = \left\{ \boldsymbol{x} \mid \boldsymbol{x} = \sum_{i\in I} \langle \boldsymbol{b}_i, \boldsymbol{x}\rangle \tilde{\boldsymbol{b}}_i, \right\} \tag{27}$$

where $I \subseteq \mathbb{N}$ is an index set (for example, in CS, I is a sparse index set such that the signal can be represented as a sparse combination of basis elements). Here, a family of functions $\{\boldsymbol{b}_i\}_{i\in\mathbb{N}}$ is usually selected as a frame that satisfies the following equality [201]:

$$A\|\boldsymbol{x}\|^2 \leq \sum_{i\in\mathbb{N}} |\langle \boldsymbol{b}_i, \boldsymbol{x}\rangle|^2 \leq B\|\boldsymbol{x}\|^2 \quad \forall \boldsymbol{x}\in H \tag{28}$$

where $A, B > 0$ are called the frame bounds, and H denotes the specific Hilbert space. If $A = B$, then the frame is called a tight frame. In (27), another family of functions $\{\tilde{\boldsymbol{b}}_i\}_{i\in\mathbb{N}}$ form the dual frame satisfying the equality

$$\sum_{i\in\mathbb{N}} \langle \boldsymbol{b}_i, \tilde{\boldsymbol{b}}_i\rangle = 1.$$

If the frame basis is chosen in a multiresolution manner, it is called *framelet*. In CS, the frames are usually chosen from wavelet transforms, overcomplete dictionaries, or other redundant bases, where the sparse combination of a subset of the frame can represent the unknown signals with high accuracy.

Even for recent approaches such as low-rank Hankel structured matrix completion approaches, the same frame interpretation exists. Specifically, for a given low-rank Hankel matrix $\mathbb{H}^{[n]}_{[d]}(\boldsymbol{x})$, let $\boldsymbol{\Phi} = [\boldsymbol{\phi}_1, \ldots, \boldsymbol{\phi}_n] \in \mathbb{R}^{n\times n}$ and $\boldsymbol{\Psi} = [\boldsymbol{\psi}_1, \ldots, \boldsymbol{\psi}_d] \in \mathbb{R}^{d\times d}$ denote the arbitrary orthonormal basis matrices that are multiplied to the left and right of the Hankel matrix, respectively. The work in [202] derived the following signal expansion, which they called the convolution framelet expansion:

$$\boldsymbol{x} \;=\; \frac{1}{d}\sum_{i=1}^{n}\sum_{j=1}^{d} \langle \boldsymbol{x}, \boldsymbol{\phi}_i \circledast \boldsymbol{\psi}_j\rangle \boldsymbol{\phi}_i \circledast \boldsymbol{\psi}_j \tag{29}$$

implying that $\{\boldsymbol{\phi}_i \circledast \boldsymbol{\psi}_j\}_{i,j=1}^{n,d}$ is a tight frame.

The observations imply that an efficient and concise signal representation is important for the success of modern image reconstruction approaches, and specific algorithms may differ in their choice of the frame basis and specific method to identify the sparse subset that concisely represents the signal.

In this perspective, the classical approaches such as CS, structured low-rank Hankel matrix approaches, etc., have two fundamental limitations. First, the choice of the underlying frame (and its dual) is based on top-down design principles. For example, most of wavelet theory has been developed around the edge-adaptive basis representations such as curvelet [203], contourlet [204], etc., whose design principle is based on top-down mathematical modeling. Moreover, the search for the sparse index set I for the case of CS is usually done using a computationally expensive optimization framework. The following section shows that these limitations of classical representation learning approaches can be largely overcome by deep learning approaches.

2) Deep Neural Networks as Combinatorial Representation Learning: The recent theory of deep convolutional framelets claims that a deep NN can be interpreted as a framelet representation, whose frame basis is learned from the training data [205]. Moreover, a recent follow-up study [206] showed how this frame representation can be automatically adapted to various input signals in a real-time manner.

To understand these findings, consider the symmetric encoder–decoder CNN in Fig. 9, which has been used for image reconstruction problems [161], [162]. Specifically, the encoder network maps a given input signal

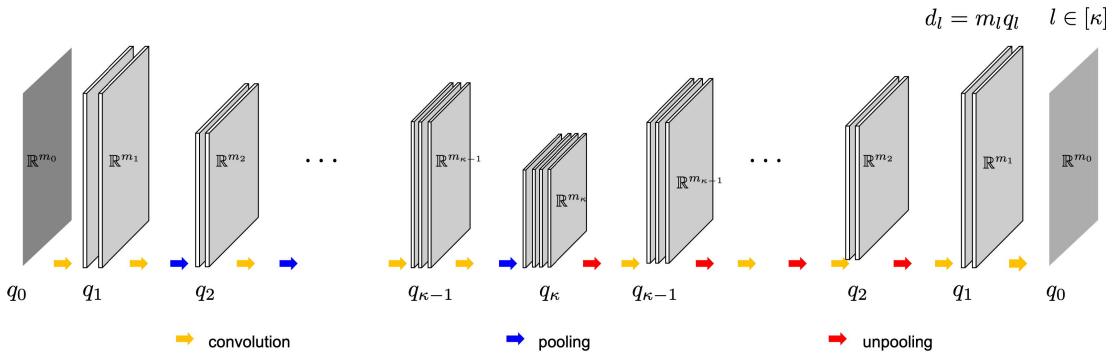

Fig. 9. *Example of encoder–decoder CNN. The encoder is composed of the first κ layers, while the latter κ layers form a decoder network.*

$\boldsymbol{x} \in \boldsymbol{\mathcal{X}} \subset \mathbb{R}^{d_0}$ to a feature space $\boldsymbol{z} \in \boldsymbol{\mathcal{Z}} \subset \mathbb{R}^{d_\kappa}$, whereas the decoder takes this feature map as an input, processes it, and produces an output $\boldsymbol{y} \in \boldsymbol{\mathcal{Y}} \subset \mathbb{R}^{d_0}$. At the lth layer, m_l, q_l, and $d_l := m_l q_l$ denote the dimension of the signal, the number of filter channels, and the total feature vector dimension, respectively. We consider a symmetric configuration, where both the encoder and decoder have the same number of layers, say κ; and the encoder layer $\mathcal{E}^l$ and the decoder layer $\mathcal{D}^l$ are symmetric

$$\begin{aligned} \mathcal{E}^l &: \mathbb{R}^{d_{l-1}} \mapsto \mathbb{R}^{d_l}, \\ \mathcal{D}^l &: \mathbb{R}^{d_l} \mapsto \mathbb{R}^{d_{l-1}}. \end{aligned}$$

The jth channel output from the the lth layer encoder can be represented by a multichannel convolution operation [206]

$$\boldsymbol{x}_j^l = \sigma\left(\boldsymbol{\Phi}^{l\top} \sum_{k=1}^{q_{l-1}} \left(\boldsymbol{x}_k^{l-1} \circledast \overline{\boldsymbol{\psi}}_{j,k}^l\right)\right) \tag{30}$$

where $\boldsymbol{x}_k^{l-1}$ denotes the kth input channel signal, $\overline{\boldsymbol{\psi}}_{j,k}^l \in \mathbb{R}^r$ denotes the r-tap convolutional kernel that is convolved with the kth input channel to contribute to the jth channel output, and $\boldsymbol{\Phi}^{l\top}$ is the pooling operator. Here, $\overline{\boldsymbol{v}}$ is the flipped version of the vector $\boldsymbol{v}$ such that $\overline{v}[n] = v[-n]$ with the periodic boundary condition, and $\circledast$ is the circular convolution (using periodic boundary conditions simplifies the mathematical treatments). Similarly, the jth channel decoder layer convolution output is given by [206]

$$\tilde{\boldsymbol{x}}_j^{l-1} = \sigma\left(\sum_{k=1}^{q_l} \left(\tilde{\boldsymbol{\Phi}}^l \tilde{\boldsymbol{x}}_k^l \circledast \tilde{\boldsymbol{\psi}}_{j,k}^l\right)\right) \tag{31}$$

where $\tilde{\boldsymbol{\Phi}}^l$ denotes the unpooling operator, $\tilde{\boldsymbol{x}}_k^l$ denotes the kth input channel signal for the decoder, and $\tilde{\boldsymbol{\psi}}_{j,k}^l \in \mathbb{R}^r$ denotes the r-tap convolutional kernel that is convolved with the kth input channel to contribute to the jth channel output.

By concatenating the multichannel signal in column direction as

$$\boldsymbol{x}^l := \left[\boldsymbol{x}_1^{l\top} \ \cdots \ \boldsymbol{x}_{q_l}^{l\top}\right]^\top$$

the encoder and decoder convolutions in (30) and (31) can be represented using the matrix notation

$$\boldsymbol{x}^l = \sigma(\boldsymbol{E}^{l\top} \boldsymbol{x}^{l-1}), \quad \tilde{\boldsymbol{x}}^{l-1} = \sigma(\boldsymbol{D}^l \tilde{\boldsymbol{x}}^l) \tag{32}$$

where $\sigma(\cdot)$ denotes the elementwise rectified linear unit (ReLU) and

$$\boldsymbol{E}^l = \begin{bmatrix} \boldsymbol{\Phi}^l \circledast \boldsymbol{\psi}_{1,1}^l & \cdots & \boldsymbol{\Phi}^l \circledast \boldsymbol{\psi}_{q_l,1}^l \\ \vdots & \ddots & \vdots \\ \boldsymbol{\Phi}^l \circledast \boldsymbol{\psi}_{1,q_{l-1}}^l & \cdots & \boldsymbol{\Phi}^l \circledast \boldsymbol{\psi}_{q_l,q_{l-1}}^l \end{bmatrix} \tag{33}$$

$$\boldsymbol{D}^l = \begin{bmatrix} \tilde{\boldsymbol{\Phi}}^l \circledast \tilde{\boldsymbol{\psi}}_{1,1}^l & \cdots & \tilde{\boldsymbol{\Phi}}^l \circledast \tilde{\boldsymbol{\psi}}_{1,q_l}^l \\ \vdots & \ddots & \vdots \\ \tilde{\boldsymbol{\Phi}}^l \circledast \tilde{\boldsymbol{\psi}}_{q_{l-1},1}^l & \cdots & \tilde{\boldsymbol{\Phi}}^l \circledast \tilde{\boldsymbol{\psi}}_{q_{l-1},q_l}^l \end{bmatrix} \tag{34}$$

and

$$\begin{aligned} \boldsymbol{\Phi}^l &= \left[\boldsymbol{\phi}_1^l \ \cdots \ \boldsymbol{\phi}_{m_l}^l\right] \\ \left[\boldsymbol{\Phi}^l \circledast \boldsymbol{\psi}_{i,j}^l\right] &:= \left[\boldsymbol{\phi}_1^l \circledast \boldsymbol{\psi}_{i,j}^l \ \cdots \ \boldsymbol{\phi}_{m_l}^l \circledast \boldsymbol{\psi}_{i,j}^l\right]. \end{aligned}$$

Then, one of the most important observations is that the output of the encoder–decoder CNN can be represented as follows [206]:

$$\boldsymbol{y} = \sum_i \langle \boldsymbol{b}_i(\boldsymbol{x}), \boldsymbol{x} \rangle \tilde{\boldsymbol{b}}_i(\boldsymbol{x}) \tag{35}$$

where $\boldsymbol{b}_i(\boldsymbol{x})$ and $\tilde{\boldsymbol{b}}_i(\boldsymbol{x})$ denote the ith columns of the following frame basis and its dual:

$$\boldsymbol{B}(\boldsymbol{x}) = \boldsymbol{E}^1 \boldsymbol{\Sigma}^1(\boldsymbol{x}) \boldsymbol{E}^2 \cdots \boldsymbol{\Sigma}^{\kappa-1}(\boldsymbol{x}) \boldsymbol{E}^\kappa \tag{36}$$

$$\tilde{\boldsymbol{B}}(\boldsymbol{x}) = \boldsymbol{D}^1 \tilde{\boldsymbol{\Sigma}}^1(\boldsymbol{x}) \boldsymbol{D}^2 \cdots \tilde{\boldsymbol{\Sigma}}^{\kappa-1}(\boldsymbol{x}) \boldsymbol{D}^\kappa \tag{37}$$

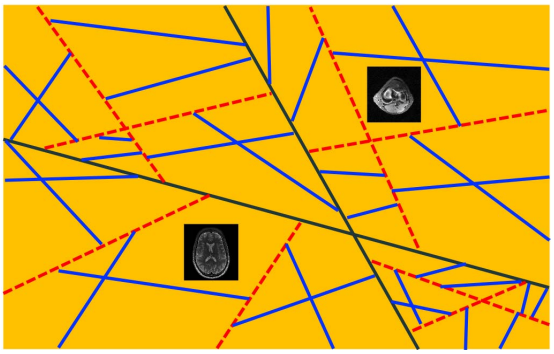

Fig. 10. *High-level illustration of input space partitioning for the three-layer NN with two filter channels for each layer. Input images at each partition share the same linear representation, but not across different partitions.*

and $\boldsymbol{\Sigma}^l(\boldsymbol{x})$ and $\tilde{\boldsymbol{\Sigma}}^l(\boldsymbol{x})$ denote diagonal matrices with 0 and 1 values that are determined by the ReLU output in the previous convolution steps. Similar basis representation holds for the encoder–decoder CNNs with the skipped connection. For more details, see [206].

In the absence of ReLU nonlinearities, the works in [207] and [208] showed that, assuming that the pooling and unpooling operators and the filter matrices satisfy appropriate frame conditions for each l, the representation (35) is indeed a frame representation of $\boldsymbol{x}$ as in (27), ensuring perfect signal reconstruction. However, in NNs, the input and output should differ, so the perfect reconstruction condition is not of practical interest. Furthermore, the signal representation in (35) should generalize well for various inputs rather than for specific inputs at the training phase.

Indeed, the work in [206] shows that the explicit dependence on the input $\boldsymbol{x}$ in (36) and (37) due to the ReLU nonlinearity solves the riddle, and CNN generalizability comes from the combinatorial nature of the expansion in (35) due to the ReLU.

Specifically, since the nonlinearity is applied after the convolution operation, the on-and-off activation pattern of each ReLU determines a binary partition of the feature space at each layer across the hyperplane that is determined by the convolution. Accordingly, in deep NNs, the input space $\mathcal{X}$ is partitioned into multiple nonoverlapping regions as shown in Fig. 10 so that input images for each region share the same linear representation but not across the partition. This implies that two different input images in Fig. 10 are automatically switched to two distinct linear representations that are different from each other.

This input adaptivity poses an important computational advantage over the classical representation learning approaches that rely on computationally expensive optimization techniques. Moreover, the representations are entirely dependent on the filter sets that are learned from the training data set, which is different from the classical representation learning approaches that are designed by mathematical principles. Furthermore, the number of input space partitions and the associated distinct linear representations increase exponentially with the network depth, width, and the skipped connection, thanks to the combinatorial nature of ReLU nonlinearities [206]. This exponentially large *expressivity* of the NN is another important advantage, which may, with the combination of the aforementioned *adaptivity*, explain the origin of the success of deep NNs for image reconstruction.

VII. OPEN QUESTIONS AND FUTURE DIRECTIONS

There are various challenges, open questions, and directions for image reconstruction that require further research. This section discusses some possible directions.

As background for this section, it is useful to recall the general topics of past (and ongoing) image reconstruction research, many of which are encapsulated in (1). The likelihood $p(\boldsymbol{y};\boldsymbol{x})$ depends on both accurate physical modeling of the imaging system and accurate statistical modeling of the measurements. Numerous papers have focused on improving the image quality by more accurate models for imaging physics or statistics. Often, more accurate models require more computation so there is an engineering challenge to make accurate models computationally tractable for routine use. Some imaging systems, like MRI, have quite flexible sampling patterns and the chosen sampling pattern is also embedded in the likelihood. The design of effective sampling patterns is an active research area in MRI. The prior $p(\boldsymbol{x})$, or the regularizer $R(\boldsymbol{x})$ depends on models for the signal $\boldsymbol{x}$. Numerous papers have focused on this aspect, and many of the open problems listed in the following also relate to signal models. The "$\arg\min$" in (1) requires an optimization algorithm, and numerous papers have proposed algorithms (often with various properties such as rapid theoretical or empirical convergence, or whose computations scale well to large-scale settings, etc.) for specific image reconstruction problems. This will continue to be an active research area because new signal models lead to new optimization problems. Furthermore, many deep-learning methods for image reconstruction are based on "unrolling" some optimization algorithm.

After performing optimization to get an image $\hat{\boldsymbol{x}}$, the research continues because the field needs task-based metrics to quantify the quality of $\hat{\boldsymbol{x}}$. Objective assessment of image quality is an ongoing research area [207] that should have more influence on image reconstruction research in the future because traditional measures like mean-squared error [208] are unlikely to predict how well a given imaging system and reconstruction method will perform in clinical tasks.

Several learning-driven iterative algorithms have been proposed for image reconstruction, particularly from limited or corrupted data and have shown promise in imaging applications. Some of these methods have proven convergence guarantees. For example, recent works [29], [126] show the convergence of BCD transform learning-based BCS algorithms to the critical points of the underlying

nonconvex problems. However, analysis of theoretical conditions on the learned models, cost functions, algorithm initializations, and (e.g., k-space) sampling guaranteeing accurate and stable image recovery in learning-based setups requires further research. Such results would shed light on appropriate model properties and constraints for different modalities and also aid the development of better behaved iterative algorithms. Theoretical results on desirable properties and invariances for filters and nonlinearities and provable ways to incorporate physics in the algorithm architecture would also benefit CNN-based reconstruction methods [34] and physics-driven deep training-based reconstruction approaches [35], [151].

In online learning-based reconstruction, adapting relatively simple models may speed up the algorithm (particularly when real-time reconstruction is needed), but at the cost of image quality and vice-versa. Developing online learning-based approaches that achieve optimal tradeoffs between complexity or richness of the learned model, runtime per minibatch, convergence (over time), and image quality is an important area of future research.

A rigorous understanding of the pros and cons of different learning-based approaches and the regimes (signal-to-noise ratios, dose levels, or undersampling) where they work well is lacking. For example, some methods learn models such as dictionaries or sparsifying transforms using model-based cost functions from the training data. These methods require fairly modest training data (e.g., several images or patches) and can learn some general properties of images that generalize fairly well [26] to new data (e.g., unseen structures may contain similar directional features). BCS methods, on the other hand, learn models on-the-fly from measurements without requiring the training data, mimicking multilayer (iterative) networks but learned in a completely unsupervised and highly adaptive manner. Supervised learning approaches learn the parameters of reconstruction models often from large data sets of input–output pairs, but may be less likely to generalize to unseen data [209] or could produce spurious reconstructions of unseen features [127] and anomalies (which are much less likely to occur in training sets). Moreover, supervised learning-based methods typically do not incorporate instance-adaptive components such as optimizing clustering for each test case within a network. A rigorous analysis of the different learning methodologies and their efficacy and drawbacks in different (training and testing) data and noise regimes would enable better use of such methods as well as aid the development of better models and improved learning-based reconstruction. Effectively and efficiently combining the benefits of both supervised and unsupervised or model-based learning methods is an interesting line of future research.

There is also increasing interest in learning-driven sampling of data, particularly limited measurements, in medical imaging. Some recent works [210], [211] proposed learning the undersampling pattern for CS MRI to minimize error in reconstructing a set of training images (e.g., prescans). The underlying optimization problems for learning the sampling were combinatorial, and moreover, the reconstruction error that is optimized would depend on the chosen reconstruction algorithm and could be a highly nonconvex function as well. These works [210], [211] proposed adapting the sampling to both the training data and the reconstruction algorithm including learning-based reconstruction schemes and showed improved image quality compared to conventional sampling strategies such as variable density random sampling for MRI. However, the learning could be computationally very expensive [211] and the convergence behavior of these sampling adaptation algorithms is unknown. Development of efficient sampling learning algorithms with guarantees would be a promising direction of future research.

Although we have discussed the recent interpretation of deep NNs with the perspective of combinatorial representation learning, there still remain open questions. For example, our understanding of the theory explains input-dependent automatic adaptation to distinct representation, but it does not prove whether the adaptation provides the optimal representation for a given input. This question is closely related to the generalization power of deep NNs, which is still an ongoing research area with many open questions. Another important open question is that even if such optimal adaptation from a filter set exists, our understanding is lacking on whether the NN training can find such optimal filter sets. Understanding the optimization landscape is another important research area in deep NNs, with many open questions [212]–[214]. Recent discoveries show that in many deep NNs, there are many global minimizers, but depending on the optimization algorithms, different minimizers are selected [215], [216]. Understanding and exploiting this "implicit bias" of the optimization algorithm is another exciting direction for future research.

Finally, given the recent trends and breakthroughs in learning for biomedical imaging, we expect that the next-generation imaging systems would leverage learning in all aspects of the imaging system. Such *smart imaging systems* may learn from big data sets (available locally in hospitals or in the cloud) as well as from real-time patient inputs and optimize the sampling for rapid (e.g., with limited measurements) or low-dose imaging and also optimize the underlying models for efficient and effective reconstruction and analytics (e.g., classification, segmentation, disease feature detection, etc.). Such adaptation of the data acquisition, reconstruction, and analytics components could be done jointly in an end-to-end manner to maximize the performance in specific clinical tasks and allowing for both radiologist and patient inputs in the learning process. The development of these next-generation learning-driven systems would involve research thrusts in both modeling and algorithmic directions coupled with innovations in physics, hardware, pulse sequence design, and so on. Importantly, we expect

models, algorithms, and computation to play an important and key role in the development of medical imaging in the near future.

VIII. CONCLUSION

This article surveyed various advances in the field of medical image reconstruction beginning with analytical approaches and simple model-based iterative reconstruction methods based on better models of the imaging system physics and sensor statistics and simple image regularization. Then, this article focused on techniques exploiting improved image models and properties such as sparsity and lowrankness that enable reconstructions from limited or corrupted data, and then discussed several recent works on sophisticated data-driven or adaptive models and machine learning techniques for reconstruction. Examples based on specific modalities, and discussions were used to provide insight into the behavior and limitations of different types of surveyed methods. We discussed different regimes of adaptivity and learning and some of the connections between different learning-based models and methods. Although the field of modeling and learning-based imaging and the concurrent interest in smart imaging systems are growing, we discussed some of the challenges, open questions, and future directions for the field in this article. ■

REFERENCES

[1] L. A. Feldkamp, L. C. Davis, and J. W. Kress, "Practical cone-beam algorithm," *J. Opt. Soc. Amer. A, Opt. Image Sci.*, vol. 1, no. 6, pp. 612–619, Jun. 1984.

[2] J. A. Fessler and B. P. Sutton, "Nonuniform fast Fourier transforms using min-max interpolation," *IEEE Trans. Signal Process.*, vol. 51, no. 2, pp. 560–574, Feb. 2003.

[3] S. De Francesco and A. M. F. da Silva, "Efficient NUFFT-based direct Fourier algorithm for fan beam CT reconstruction," *Proc. SPIE*, vol. 5370, pp. 666–677, May 2004.

[4] K. Sauer and C. Bouman, "A local update strategy for iterative reconstruction from projections," *IEEE Trans. Signal Process.*, vol. 41, no. 2, pp. 534–548, Feb. 1993.

[5] J.-B. Thibault, C. A. Bouman, K. D. Sauer, and J. Hsieh, "A recursive filter for noise reduction in statistical iterative tomographic imaging," *Proc. SPIE*, vol. 6065, Feb. 2006, Art. no. 60650X.

[6] A. R. De Pierro, "On the relation between the ISRA and the EM algorithm for positron emission tomography," *IEEE Trans. Med. Imag.*, vol. 12, no. 2, pp. 328–333, Jun. 1993.

[7] K. P. Pruessmann, "Encoding and reconstruction in parallel MRI," *NMR Biomed.*, vol. 19, no. 3, pp. 288–299, May 2006.

[8] P. Feng and Y. Bresler, "Spectrum-blind minimum-rate sampling and reconstruction of multiband signals," in *Proc. ICASSP*, vol. 3, May 1996, pp. 1688–1691.

[9] Y. Bresler and P. Feng, "Spectrum-blind minimum-rate sampling and reconstruction of 2-D multiband signals," in *Proc. 3rd IEEE Int. Conf. Image Process. (ICIP)*, Sep. 1996, pp. 701–704.

[10] D. L. Donoho, "Compressed sensing," *IEEE Trans. Inf. Theory*, vol. 52, no. 4, pp. 1289–1306, Apr. 2006.

[11] E. J. Candès, J. Romberg, and T. Tao, "Robust uncertainty principles: Exact signal reconstruction from highly incomplete frequency information," *IEEE Trans. Inf. Theory*, vol. 52, no. 2, pp. 489–509, Feb. 2006.

[12] J. C. Ye, Y. Bresler, and P. Moulin, "A self-referencing level-set method for image reconstruction from sparse Fourier samples," *Int. J. Comput. Vis.*, vol. 50, no. 3, pp. 253–270, 2002.

[13] R. G. Baraniuk, E. J. Candès, M. Elad, and Y. Ma, "Applications of sparse representation and compressive sensing," *Proc. IEEE*, vol. 98, no. 6, pp. 906–909, Jun. 2010.

[14] G. Wang, Y. Bresler, and V. Ntziachristos, "Guest editorial compressive sensing for biomedical imaging," *IEEE Trans. Med. Imag.*, vol. 30, no. 5, pp. 1013–1016, May 2011.

[15] 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.

[16] M. Lustig, D. L. Donoho, J. M. Santos, and J. M. Pauly, "Compressed sensing MRI," *IEEE Signal Process. Mag.*, vol. 25, no. 2, pp. 72–82, Mar. 2008.

[17] FDA. (2017). *510k Premarket Notification of HyperSense (GE Medical Systems)*. [Online]. Available: https://www.accessdata.fda.gov/scripts/cdrh/cfdocs/cfpmn/pmn.cfm?ID=K162722

[18] FDA. (2017). *510k Premarket Notification of Compressed Sensing Cardiac Cine (Siemens)*. [Online]. Available: https://www.accessdata.fda.gov/scripts/cdrh/cfdocs/cfpmn/pmn.cfm?ID=K163312

[19] FDA. (2018). *510k Premarket Notification of Compressed SENSE*. [Online]. Available: https://www.accessdata.fda.gov/cdrh_docs/pdf17/K173079.pdf

[20] X. Qu, W. Zhang, D. Guo, C. Cai, S. Cai, and Z. Chen, "Iterative thresholding compressed sensing MRI based on contourlet transform," *Inverse Problems Sci. Eng.*, vol. 18, no. 6, pp. 737–758, 2010.

[21] B. Adcock, A. C. Hansen, C. Poon, and B. Roman, "Breaking the coherence barrier: A new theory for compressed sensing," 2013, *arXiv:1302.0561*. [Online]. Available: https://arxiv.org/abs/1302.0561

[22] D. Kim, S. Ramani, and J. A. Fessler, "Combining ordered subsets and momentum for accelerated X-ray CT image reconstruction," *IEEE Trans. Med. Imag.*, vol. 34, no. 1, pp. 167–178, Jan. 2015.

[23] M. Aharon, M. Elad, and A. Bruckstein, "K-SVD: An algorithm for designing overcomplete dictionaries for sparse representation," *IEEE Trans. Signal Process.*, vol. 54, no. 11, pp. 4311–4322, Oct. 2006.

[24] S. Ravishankar and Y. Bresler, "Learning sparsifying transforms," *IEEE Trans. Signal Process.*, vol. 61, no. 5, pp. 1072–1086, Mar. 2013.

[25] Q. Xu, H. Yu, X. Mou, L. Zhang, J. Hsieh, and G. Wang, "Low-dose X-ray CT reconstruction via dictionary learning," *IEEE Trans. Med. Imag.*, vol. 31, no. 9, pp. 1682–1697, Sep. 2012.

[26] X. Zheng, S. Ravishankar, Y. Long, and J. A. Fessler, "PWLS-ULTRA: An efficient clustering and learning-based approach for low-dose 3D CT image reconstruction," *IEEE Trans. Med. Imag.*, vol. 37, no. 6, pp. 1498–1510, Jun. 2018.

[27] S. Ravishankar and Y. Bresler, "MR image reconstruction from highly undersampled k-space data by dictionary learning," *IEEE Trans. Med. Imag.*, vol. 30, no. 5, pp. 1028–1041, May 2011.

[28] S. G. Lingala and M. Jacob, "Blind compressive sensing dynamic MRI," *IEEE Trans. Med. Imag.*, vol. 32, no. 6, pp. 1132–1145, Jun. 2013.

[29] S. Ravishankar and Y. Bresler, "Data-driven learning of a union of sparsifying transforms model for blind compressed sensing," *IEEE Trans. Comput. Imag.*, vol. 2, no. 3, pp. 294–309, May 2016.

[30] S. Gleichman and Y. C. Eldar, "Blind compressed sensing," *IEEE Trans. Inf. Theory*, vol. 57, no. 10, pp. 6958–6975, Oct. 2011.

[31] M. Mardani, G. Mateos, and G. B. Giannakis, "Subspace learning and imputation for streaming big data matrices and tensors," *IEEE Trans. Signal Process.*, vol. 63, no. 10, pp. 2663–2677, May 2015.

[32] B. E. Moore, S. Ravishankar, R. R. Nadakuditi, and J. A. Fessler, "Online adaptive image reconstruction (OnAIR) using dictionary models," *IEEE Trans. Comput. Imag.*, 2019. [Online]. Available: https://arxiv.org/abs/1809.01817

[33] 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, Feb. 2018.

[34] D. Lee, J. Yoo, S. Tak, and J. C. Ye, "Deep residual learning for accelerated MRI using magnitude and phase networks," *IEEE Trans. Biomed. Eng.*, vol. 65, no. 9, pp. 1985–1995, Sep. 2018.

[35] S. Ravishankar, I. Y. Chun, and J. A. Fessler, "Physics-driven deep training of dictionary-based algorithms for MR image reconstruction," in *Proc. 51st Asilomar Conf. Signals, Syst., Comput.*, 2017, pp. 1859–1863.

[36] G. Wang, "A perspective on deep imaging," *IEEE Access*, vol. 4, pp. 8914–8924, 2016.

[37] H. Shan *et al.*, "Competitive performance of a modularized deep neural network compared to commercial algorithms for low-dose CT image reconstruction," *Nature Mach. Intell.*, vol. 1, no. 6, pp. 269–276, 2019.

[38] G. Wang, J. C. Ye, K. Mueller, and J. A. Fessler, "Image reconstruction is a new frontier of machine learning," *IEEE Trans. Med. Imag.*, vol. 37, no. 6, pp. 1289–1296, Jun. 2018.

[39] C. E. Floyd, "An artificial neural network for SPECT image reconstruction," *IEEE Trans. Med. Imag.*, vol. 10, no. 3, pp. 485–487, Sep. 1991.

[40] S. Venkataraman, Z.-P. Liang, and R. L. Magin, "A neural network approach to NMR spectral estimation," in *Proc. Int. Soc. Mag. Res. Med.*, 1994, p. 1214.

[41] K. P. Pruessmann, M. Weiger, P. Börnert, and P. Boesiger, "Advances in sensitivity encoding with arbitrary k-space trajectories," *Mag. Res. Med.*, vol. 46, pp. 638–651, Oct. 2001.

[42] S. Geman and D. Geman, "Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. PMAI-6, no. 6, pp. 721–741, Nov. 1984.

[43] J. Besag, "On the statistical analysis of dirty pictures," *J. Roy. Stat. Soc. B*, vol. 48, no. 3, pp. 259–302, 1986.

[44] C.-M. Kao, X. Pan, C.-T. Chen, and W. H. Wong, "Image restoration and reconstruction with a Bayesian approach," *Med. Phys.*, vol. 25, pp. 600–613, May 1998.

[45] C. T. Chen, X. Ouyang, W. H. Wong, X. Hu, V. E. Johnson, C. Ordonez, and C. Metz, "Sensor fusion in image reconstruction," *IEEE Trans. Nucl. Sci.*, vol. 38, no. 2, pp. 687–692, Apr. 1991.

[46] G. Gindi, M. Lee, A. Rangarajan, and I. G. Zubal, "Bayesian reconstruction of functional images using anatomical information as priors," *IEEE Trans. Med. Imag.*, vol. 12, no. 4, pp. 670–680, Dec. 1993.

[47] Y. Cao and D. N. Levin, "Using prior knowledge of human anatomy to constrain MR image acquisition and reconstruction: Half k-space and full k-space techniques," *Magn. Reson. Imag.*, vol. 15, no. 6, pp. 669–677, 1997.

[48] Z.-P. Liang and P. C. Lauterbur, "A generalized series approach to MR spectroscopic imaging," *IEEE Trans. Med. Imag.*, vol. 10, no. 2, pp. 132–137, Jun. 1991.

[49] H. M. Hudson and R. S. Larkin, "Accelerated image reconstruction using ordered subsets of projection data," *IEEE Trans. Med. Imag.*, vol. 13, no. 4, pp. 601–609, Dec. 1994.

[50] S. Ahn *et al.*, "Quantitative comparison of OSEM and penalized likelihood image reconstruction using relative difference penalties for clinical PET," *Phys. Med. Biol.*, vol. 60, pp. 5733–5752, Aug. 2015.

[51] C. Bouman and K. Sauer, "A generalized Gaussian image model for edge-preserving MAP estimation," *IEEE Trans. Image Process.*, vol. 2, no. 3, pp. 296–310, Jul. 1993.

[52] J. A. Fessler and W. L. Rogers, "Spatial resolution properties of penalized-likelihood image reconstruction: Space-invariant tomographs," *IEEE Trans. Image Process.*, vol. 5, no. 9, pp. 1346–1358, Sep. 1996.

[53] J. Nuyts, D. Beque, P. Dupont, and L. Mortelmans, "A concave prior penalizing relative differences for maximum-a-posteriori reconstruction in emission tomography," *IEEE Trans. Nucl. Sci.*, vol. 49, no. 1, pp. 56–60, Feb. 2002.

[54] J.-B. Thibault, K. Sauer, C. Bouman, and J. Hsieh, "A three-dimensional statistical approach to improved image quality for multi-slice helical CT," *Med. Phys.*, vol. 34, pp. 4526–4544, Nov. 2007.

[55] R. M. Leahy and B. D. Jeffs, "On the design of maximally sparse beamforming arrays," *IEEE Trans. Antennas Propag.*, vol. 39, no. 8, pp. 1178–1187, Aug. 1991.

[56] G. Harikumar and Y. Bresler, "A new algorithm for computing sparse solutions to linear inverse problems," in *Proc. IEEE Conf. Acoust. Speech Signal Process.*, vol. 3, May 1996, pp. 1331–1334.

[57] Z.-P. Liang, E. M. Haacke, and C. W. Thomas, "High-resolution inversion of finite Fourier transform data through a localised polynomial approximation," *Inverse Problems*, vol. 5, pp. 831–848, Oct. 1989.

[58] R. Chartrand, "Fast algorithms for nonconvex compressive sensing: MRI reconstruction from very few data," in *Proc. IEEE Int. Symp. Biomed. Imag., From Nano Macro*, Jun. 2009, pp. 262–265.

[59] L. Geerts-Ossevoort *et al.*, "Compressed SENSE," Philips, Amsterdam, The Netherlands, White Paper 4522 991 31821, Nov. 2018.

[60] G.-H. Chen, J. Tang, and S. Leng, "Prior image constrained compressed sensing (PICCS): A method to accurately reconstruct dynamic CT images from highly undersampled projection data sets," *Med. Phys.*, vol. 35, no. 2, pp. 660–663, Feb. 2008.

[61] Y. Liu *et al.*, "Balanced sparse model for tight frames in compressed sensing magnetic resonance imaging," *PLoS ONE*, vol. 10, pp. 1–19, Apr. 2015.

[62] Y. Liu, Z. Zhan, J.-F. Cai, D. Guo, Z. Chen, and X. Qu, "Projected iterative soft-thresholding algorithm for tight frames in compressed sensing magnetic resonance imaging," *IEEE Trans. Med. Imag.*, vol. 35, no. 9, pp. 2130–2140, Sep. 2016.

[63] J. P. Haldar, "Low-rank modeling of local-space neighborhoods (LORAKS) for constrained MRI," *IEEE Trans. Med. Imag.*, vol. 33, no. 3, pp. 668–681, Mar. 2014.

[64] K. H. Jin, D. Lee, and J. C. Ye, "A general framework for compressed sensing and parallel MRI using annihilating filter based low-rank Hankel matrix," *IEEE Trans. Comput. Imag.*, vol. 2, no. 4, pp. 480–495, Dec. 2016.

[65] D. Lee, K. H. Jin, E. Y. Kim, S.-H. Park, and J. C. Ye, "Acceleration of MR parameter mapping using annihilating filter-based low rank Hankel matrix (ALOHA)," *Magn. Reson. Med.*, vol. 76, no. 6, pp. 1848–1864, 2016.

[66] G. Ongie and M. Jacob, "Off-the-grid recovery of piecewise constant images from few Fourier samples," *SIAM J. Imag. Sci.*, vol. 9, no. 3, pp. 1004–1041, 2015.

[67] G. Ongie and M. Jacob, "A fast algorithm for convolutional structured low-rank matrix recovery," *IEEE Trans. Comput. Imag.*, vol. 3, no. 4, pp. 535–550, Dec. 2017.

[68] K. H. Jin, J.-Y. Um, D. Lee, J. Lee, S.-H. Park, and J. C. Ye, "MRI artifact correction using sparse + low-rank decomposition of annihilating filter-based Hankel matrix," *Magn. Reson. Med.*, vol. 78, no. 1, pp. 327–340, 2017.

[69] V. Chandrasekaran, B. Recht, P. A. Parrilo, and A. S. Willsky, "The convex geometry of linear inverse problems," *Found. Comput. Math.*, vol. 12, no. 6, pp. 805–849, Dec. 2012.

[70] Z. P. Liang, "Spatiotemporal imaging with partially separable functions," in *Proc. IEEE Int. Symp. Biomed. Imag., Nano Macro*, Apr. 2007, pp. 988–991.

[71] V. Singh, A. H. Tewfik, and D. B. Ress, "Under-sampled functional MRI using low-rank plus sparse matrix decomposition," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP)*, Apr. 2015, pp. 897–901.

[72] M. Chiew, N. N. Graedel, J. A. McNab, S. M. Smith, and K. L. Miller, "Accelerating functional MRI using fixed-rank approximations and radial-Cartesian sampling," *Magn. Reson. Med.*, vol. 76, no. 6, pp. 1825–1836, 2016.

[73] Y. Hu *et al.*, "Motion-robust reconstruction of multishot diffusion-weighted images without phase estimation through locally low-rank regularization," *Magn. Reson. Med.*, vol. 81, no. 2, pp. 1181–1190, 2019.

[74] G. Mazor, L. Weizman, A. Tal, and Y. C. Eldar, "Low-rank magnetic resonance fingerprinting," *Med. Phys.*, vol. 45, no. 9, pp. 4066–4084, Sep. 2018.

[75] B. Zhao *et al.*, "Improved magnetic resonance fingerprinting reconstruction with low-rank and subspace modeling," *Magn. Reson. Med.*, vol. 79, no. 2, pp. 933–942, 2018.

[76] G. L. da Cruz, A. Bustin, O. Jaubert, T. Schneider, R. M. Botnar, and C. Prieto, "Sparsity and locally low rank regularization for MR fingerprinting," *Magn. Reson. Med.*, vol. 81, no. 6, pp. 3530–3543, 2019.

[77] J. P. Haldar and Z. P. Liang, "Spatiotemporal imaging with partially separable functions: A matrix recovery approach," in *Proc. IEEE Int. Symp. Biomed. Imag., Nano Macro*, Apr. 2010, pp. 716–719.

[78] B. Zhao, J. P. Haldar, C. Brinegar, and Z.-P. Liang, "Low rank matrix recovery for real-time cardiac MRI," in *Proc. IEEE Int. Symp. Biomed. Imag., Nano Macro*, Apr. 2010, pp. 996–999.

[79] H. Pedersen, S. Kozerke, S. Ringgaard, K. Nehrke, and W. Y. Kim, "k-t PCA: Temporally constrained k-t BLAST reconstruction using principal component analysis," *Magn. Reson. Med.*, vol. 62, no. 3, pp. 706–716, 2009.

[80] J. Trzasko and A. Manduca, "Local versus global low-rank promotion in dynamic MRI series reconstruction," in *Proc. ISMRM*, 2011, p. 4371.

[81] 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, May 2011.

[82] B. Zhao, J. P. Haldar, A. G. Christodoulou, and Z. P. Liang, "Image reconstruction from highly undersampled (k, t)-space data with joint partial separability and sparsity constraints," *IEEE Trans. Med. Imag.*, vol. 31, no. 9, pp. 1809–1820, Sep. 2012.

[83] E. J. Candès, X. Li, Y. Ma, and J. Wright, "Robust principal component analysis?" *J. ACM*, vol. 58, no. 3, pp. 11:1–11:37, May 2011, Art. no. 11.

[84] H. Guo, C. Qiu, and N. Vaswani, "An online algorithm for separating sparse and low-dimensional signal sequences from their sum," *IEEE Trans. Signal Process.*, vol. 62, no. 16, pp. 4284–4297, Aug. 2014.

[85] R. Otazo, E. J. Candès, and D. K. Sodickson, "Low-rank plus sparse matrix decomposition for accelerated dynamic MRI with separation of background and dynamic components," *Magn. Reson. Med.*, vol. 73, no. 3, pp. 1125–1136, Mar. 2015.

[86] B. Trémoulhéac, N. Dikaios, D. Atkinson, and S. R. Arridge, "Dynamic MR image reconstruction–separation from undersampled (k, t)-space via low-rank plus sparse prior," *IEEE Trans. Med. Imag.*, vol. 33, no. 8, pp. 1689–1701, Aug. 2014.

[87] D. Banco, S. Aeron, and W. S. Hoge, "Sampling and recovery of MRI data using low rank tensor models," in *Proc. 38th Annu. Int. Conf. IEEE Eng. Med. Biol. Soc. (EMBC)*, Aug. 2016, pp. 448–452.

[88] B. Yaman, S. Weingärtner, N. Kargas, N. D. Sidiropoulos, and M. Akcakaya, "Locally low-rank tensor regularization for high-resolution quantitative dynamic MRI," in *Proc. IEEE 7th Int. Workshop Comput. Adv. Multi-Sensor Adapt. Process. (CAMSAP)*, Dec. 2017, pp. 1–5.

[89] J. He, Q. Liu, A. G. Christodoulou, C. Ma, F. Lam, and Z.-P. Liang, "Accelerated high-dimensional MR imaging with sparse sampling using low-rank tensors," *IEEE Trans. Med. Imag.*, vol. 35, no. 9, pp. 2119–2129, Sep. 2016.

[90] A. G. Christodoulou *et al.*, "Magnetic resonance multitasking for motion-resolved quantitative cardiovascular imaging," *Nature Biomed. Eng.*, vol. 2, no. 4, pp. 215–226, 2018.

[91] J. Lee, K. H. Jin, and J. C. Ye, "Reference-free single-pass EPI Nyquist ghost correction using annihilating filter-based low rank Hankel matrix (ALOHA)," *Magn. Reson. Med.*, vol. 76, no. 6, pp. 1775–1789, 2016.

[92] J. C. Ye, J. M. Kim, K. H. Jin, and K. Lee, "Compressive sampling using annihilating filter-based low-rank interpolation," *IEEE Trans. Inf. Theory*, vol. 63, no. 2, pp. 777–801, Feb. 2017.

[93] G. Ongie, S. Biswas, and M. Jacob, "Convex recovery of continuous domain piecewise constant images from nonuniform Fourier samples," *IEEE Trans. Signal Process.*, vol. 66, no. 1, pp. 236–250, Jan. 2017.

[94] M. Vetterli, P. Marziliano, and T. Blu, "Sampling signals with finite rate of innovation," *IEEE Trans. Signal Process.*, vol. 50, no. 6, pp. 1417–1428, Jun. 2002.

[95] I. Maravic and M. Vetterli, "Sampling and reconstruction of signals with finite rate of innovation in the presence of noise," *IEEE Trans. Signal Process.*, vol. 53, no. 8, pp. 2788–2805, Aug. 2005.

[96] E. M. Haacke, Z.-P. Liang, and S. H. Izen, "Superresolution reconstruction through object modeling and parameter estimation," *IEEE Trans. Acoust., Speech Signal Process.*, vol. 37, no. 4, pp. 592–595, Apr. 1989.

[97] P. J. Shin *et al.*, "Calibrationless parallel imaging reconstruction based on structured low-rank matrix completion," *Magn. Reson. Med.*, vol. 72, no. 4, pp. 959–970, 2014.

[98] J. Min, K. H. Jin, M. Unser, and J. C. Ye, "Grid-free localization algorithm using low-rank Hankel matrix for super-resolution microscopy," *IEEE Trans. Image Process.*, vol. 27, no. 10, pp. 4771–4786, Oct. 2018.

[99] K. H. Jin and J. C. Ye, "Annihilating filter-based low-rank Hankel matrix approach for image inpainting," *IEEE Trans. Image Process.*, vol. 24, no. 11, pp. 3498–3511, Nov. 2015.

[100] K. H. Jin and J. C. Ye, "Sparse and low-rank decomposition of a Hankel structured matrix for impulse noise removal," *IEEE Trans. Image Process.*, vol. 27, no. 3, pp. 1448–1461, Mar. 2018.

[101] X. Qu *et al.*, "Undersampled MRI reconstruction with patch-based directional wavelets," *Magn.*

Reson. Imag., vol. 30, no. 7, pp. 964–977, Sep. 2012.
[102] B. Ning, X. Qu, D. Guo, C. Hu, and Z. Chen, "Magnetic resonance image reconstruction using trained geometric directions in 2D redundant wavelets domain and non-convex optimization," *Magn. Reson. Imag.*, vol. 31, no. 9, pp. 1611–1622, 2013.
[103] Z. Zhan, J.-F. Cai, D. Guo, Y. Liu, Z. Chen, and X. Qu, "Fast multiclass dictionaries learning with geometrical directions in MRI reconstruction," *IEEE Trans. Biomed. Eng.*, vol. 63, no. 9, pp. 1850–1861, Sep. 2016.
[104] X. Qu, Y. Hou, F. Lam, D. Guo, J. Zhong, and Z. Chen, "Magnetic resonance image reconstruction from undersampled measurements using a patch-based nonlocal operator," *Med. Image Anal.*, vol. 18, pp. 843–856, Aug. 2014.
[105] M. Akcakaya *et al.*, "Low-dimensional-structure self-learning and thresholding: Regularization beyond compressed sensing for MRI reconstruction," *Mag. Res. Med.*, vol. 66, no. 3, pp. 756–767, 2011.
[106] R. Vidal, "Subspace clustering," *IEEE Signal Process. Mag.*, vol. 28, no. 2, pp. 52–68, Mar. 2011.
[107] B. A. Olshausen and D. J. Field, "Emergence of simple-cell receptive field properties by learning a sparse code for natural images," *Nature*, vol. 381, no. 6583, pp. 607–609, 1996.
[108] J. Mairal, F. Bach, J. Ponce, and G. Sapiro, "Online learning for matrix factorization and sparse coding," *J. Mach. Learn. Res.*, vol. 11, pp. 19–60, Mar. 2010.
[109] M. Elad and M. Aharon, "Image denoising via sparse and redundant representations over learned dictionaries," *IEEE Trans. Image Process.*, vol. 15, no. 12, pp. 3736–3745, Dec. 2006.
[110] J. Mairal, M. Elad, and G. Sapiro, "Sparse representation for color image restoration," *IEEE Trans. Image Process.*, vol. 17, no. 1, pp. 53–69, Jan. 2008.
[111] A. M. Bruckstein, D. L. Donoho, and M. Elad, "From sparse solutions of systems of equations to sparse modeling of signals and images," *SIAM Rev.*, vol. 51, no. 1, pp. 34–81, 2009.
[112] R. Rubinstein, M. Zibulevsky, and M. Elad, "Double sparsity: Learning sparse dictionaries for sparse signal approximation," *IEEE Trans. Signal Process.*, vol. 58, no. 3, pp. 1553–1564, Mar. 2010.
[113] Y. Pati, R. Rezaiifar, and P. Krishnaprasad, "Orthogonal matching pursuit: Recursive function approximation with applications to wavelet decomposition," in *Proc. Asilomar Conf. Signals, Syst. Comput.*, vol. 1, 1993, pp. 40–44.
[114] Y. Wang and L. Ying, "Compressed sensing dynamic cardiac cine MRI using learned spatiotemporal dictionary," *IEEE Trans. Biomed. Eng.*, vol. 61, no. 4, pp. 1109–1120, Apr. 2014.
[115] J. Caballero, A. N. Price, D. Rueckert, and J. V. Hajnal, "Dictionary learning and time sparsity for dynamic MR data reconstruction," *IEEE Trans. Med. Imag.*, vol. 33, no. 4, pp. 979–994, Apr. 2014.
[116] D. Weller, "Reconstruction with dictionary learning for accelerated parallel magnetic resonance imaging," in *Proc. IEEE Southwest Symp. Image Anal. Interpretation (SSIAI)*, Mar. 2016, pp. 105–108.
[117] S. Chen, H. Liu, P. Shi, and Y. Chen, "Sparse representation and dictionary learning penalized image reconstruction for positron emission tomography," *Phys. Med. Biol.*, vol. 60, no. 2, pp. 807–823, Jan. 2015.
[118] Y. Huang, J. Paisley, Q. Lin, X. Ding, X. Fu, and X.-P. Zhang, "Bayesian nonparametric dictionary learning for compressed sensing MRI," *IEEE Trans. Image Process.*, vol. 23, no. 12, pp. 5007–5019, Dec. 2014.
[119] S. Tan *et al.*, "Tensor-based dictionary learning for dynamic tomographic reconstruction," *Phys. Med. Biol.*, vol. 60, no. 7, pp. 2803–2818, Apr. 2015.
[120] Y. Zhang, X. Mou, G. Wang, and H. Yu, "Tensor-based dictionary learning for spectral CT reconstruction," *IEEE Trans. Med. Imag.*, vol. 36, no. 1, pp. 142–154, Jan. 2017.
[121] S. Ravishankar, R. R. Nadakuditi, and J. A. Fessler, "Efficient sum of outer products dictionary learning (SOUP-DIL) and its application to inverse problems," *IEEE Trans. Comput. Imag.*, vol. 3, no. 4, pp. 694–709, Dec. 2017.
[122] S. Ravishankar, B. E. Moore, R. R. Nadakuditi, and J. A. Fessler, "Low-rank and adaptive sparse signal (LASSI) models for highly accelerated dynamic imaging," *IEEE Trans. Med. Imag.*, vol. 36, no. 5, pp. 1116–1128, May 2017.
[123] C. Garcia-Cardona and B. Wohlberg, "Convolutional dictionary learning: A comparative review and new algorithms," *IEEE Trans. Comput. Imag.*, vol. 4, no. 3, pp. 366–381, Sep. 2018.
[124] B. Wohlberg, "Efficient algorithms for convolutional sparse representations," *IEEE Trans. Image Process.*, vol. 25, no. 1, pp. 301–315, Jan. 2016.
[125] I. Y. Chun and J. A. Fessler, "Convolutional dictionary learning: Acceleration and convergence," *IEEE Trans. Image Process.*, vol. 27, no. 4, pp. 1697–1712, Apr. 2018.
[126] S. Ravishankar and Y. Bresler, "Efficient blind compressed sensing using sparsifying transforms with convergence guarantees and application to magnetic resonance imaging," *SIAM J. Imag. Sci.*, vol. 8, no. 4, pp. 2519–2557, 2015.
[127] B. Wen, S. Ravishankar, L. Pfister, and Y. Bresler, "Transform learning for magnetic resonance image reconstruction: From model-based learning to building neural networks," 2019, *arXiv:1903.11431*. [Online]. Available: https://arxiv.org/abs/1903.11431
[128] M. W. Marcellin, M. J. Gormish, A. Bilgin, and M. P. Boliek, "An overview of JPEG-2000," in *Proc. Data Compress. Conf.*, 2000, pp. 523–541.
[129] A. K. Tanc and E. M. Eksioglu, "MRI reconstruction with joint global regularization and transform learning," *Comput. Med. Imag. Graph.*, vol. 53, pp. 1–8, Oct. 2016.
[130] L. Pfister and Y. Bresler, "Model-based iterative tomographic reconstruction with adaptive sparsifying transforms," *Proc. SPIE*, vol. 9020, pp. 90200H-1–90200H-11, Mar. 2014.
[131] S. Ye, S. Ravishankar, Y. Long, and J. A. Fessler, "Adaptive sparse modeling and shifted-Poisson likelihood based approach for low-dose CT image reconstruction," in *Proc. IEEE 27th Int. Workshop Mach. Learn. Signal Process. (MLSP)*, Sep. 2017, pp. 1–6.
[132] S. Ravishankar and Y. Bresler, "Learning overcomplete sparsifying transforms for signal processing," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process.*, May 2013, pp. 3088–3092.
[133] B. Wen, S. Ravishankar, and Y. Bresler, "Structured overcomplete sparsifying transform learning with convergence guarantees and applications," *Int. J. Comput. Vis.*, vol. 114, nos. 2–3, pp. 137–167, 2015.
[134] W. P. Segars, M. Mahesh, T. J. Beck, E. C. Frey, and B. M. W. Tsui, "Realistic CT simulation using the 4D XCAT phantom," *Med. Phys.*, vol. 35, pp. 3800–3808, Aug. 2008.
[135] J. H. Cho and J. A. Fessler, "Regularization designs for uniform spatial resolution and noise properties in statistical image reconstruction for 3-D X-ray CT," *IEEE Trans. Med. Imag.*, vol. 34, no. 2, pp. 678–689, Feb. 2015.
[136] Z. Li, S. Ravishankar, Y. Long, and J. A. Fessler, "DECT-MULTRA: Dual-energy CT image decomposition with learned mixed material models and efficient clustering," 2019, *arXiv:1901.00106*. [Online]. Available: https://arxiv.org/abs/1901.00106
[137] Z. Li, S. Ravishankar, and Y. Long, "Image-domain multi-material decomposition using a union of cross-material models," *Proc. SPIE*, vol. 11072, pp. 1107210-1–1107210-5, May 2019.
[138] B. Wen, S. Ravishankar, and Y. Bresler, "FRIST—Flipping and rotation invariant sparsifying transform learning and applications," *Inverse Problems*, vol. 33, no. 7, 2017, Art. no. 074007.
[139] B. Wen, Y. Li, and Y. Bresler, "The power of complementary regularizers: Image recovery via transform learning and low-rank modeling," 2018, *arXiv:1808.01316*. [Online]. Available: https://arxiv.org/abs/1808.01316
[140] K. Dabov, A. Foi, V. Katkovnik, and K. Egiazarian, "Image denoising by sparse 3D transform-domain collaborative filtering," *IEEE Trans. Image Process.*, vol. 16, no. 8, pp. 2080–2095, Aug. 2007.
[141] Y. Yang, J. Sun, H. Li, and Z. Xu, "Deep ADMM-net for compressive sensing MRI," in *Proc. Adv. Neural Inf. Process. Syst.*, 2016, pp. 10–18.
[142] S. Ravishankar, B. E. Moore, R. R. Nadakuditi, and J. A. Fessler, "Efficient online dictionary adaptation and image reconstruction for dynamic MRI," in *Proc. 51st Asilomar Conf. Signals, Syst., Comput.*, 2017, pp. 835–839.
[143] M. Mardani, G. B. Giannakis, and K. Ugurbil, "Tracking tensor subspaces with informative random sampling for real-time MR imaging," 2016, *arXiv:1609.04104*. [Online]. Available: https://arxiv.org/abs/1609.04104
[144] L. Pfister and Y. Bresler, "Learning filter bank sparsifying transforms," *IEEE Trans. Signal Process.*, vol. 67, no. 2, pp. 504–519, Jan. 2019.
[145] I. Y. Chun and J. A. Fessler, "Convolutional analysis operator learning: Application to sparse-view CT," in *Proc. IEEE Asilomar Conf. Signals, Syst., Comput.*, Oct. 2018, pp. 1631–1635.
[146] I. Y. Chun and J. A. Fessler, "Convolutional analysis operator learning: Acceleration and convergence," 2018, *arXiv:1802.05584*. [Online]. Available: https://arxiv.org/abs/1802.05584
[147] S. Ravishankar, A. Lahiri, C. Blocker, and J. A. Fessler, "Deep dictionary-transform learning for image reconstruction," in *Proc. IEEE 15th Int. Symp. Biomed. Imag. (ISBI)*, Feb. 2018, pp. 1208–1212.
[148] S. Ravishankar and B. Wohlberg, "Learning multi-layer transform models," in *Proc. 56th Annu. Allerton Conf. Commun., Control, Comput. (Allerton)*, 2018, pp. 160–165.
[149] L. Pfister and Y. Bresler, "Learning sparsifying filter banks," *Proc. SPIE*, vol. 9597, pp. 959703-1–959703-10, Aug. 2015.
[150] X. Zheng, S. Ravishankar, Y. Long, M. L. Klasky, and B. Wohlberg, "Multi-layer residual sparsifying transform learning for image reconstruction," 2019, *arXiv:1906.00165*. [Online]. Available: https://arxiv.org/abs/1906.00165
[151] Y. Chun and J. A. Fessler, "Deep BCD-Net using identical encoding-decoding CNN structures for iterative image recovery," in *Proc. IEEE 13th Image, Video, Multidimensional Signal Process. Workshop (IVMSP)*, Jun. 2018, pp. 1–5.
[152] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet classification with deep convolutional neural networks," in *Proc. Adv. Neural Inf. Process. Syst.*, 2012, pp. 1097–1105.
[153] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2016, pp. 770–778.
[154] O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," in *Proc. 18th Int. Conf. Med. Image Comput.-Assist. Intervent*, Munich, Germany, 2015, pp. 234–241.
[155] K. Zhang, W. Zuo, Y. Chen, D. Meng, and L. Zhang, "Beyond a Gaussian Denoiser: Residual learning of deep CNN for image denoising," *IEEE Trans. Image Process.*, vol. 26, no. 7, pp. 3142–3155, Jul. 2017.
[156] E. Kang, J. Min, and J. C. Ye, "A deep convolutional neural network using directional wavelets for low-dose X-ray CT reconstruction," *Med. Phys.*, vol. 44, no. 10, pp. e360–e375, 2017.
[157] H. Chen *et al.*, "Low-dose CT with a residual encoder-decoder convolutional neural network," *IEEE Trans. Image Process.*, vol. 36, no. 12, pp. 2524–2535, Dec. 2017.
[158] E. Kang, W. Chang, J. Yoo, and J. C. Ye, "Deep convolutional framelet denosing for low-dose CT via wavelet residual network," *IEEE Trans. Med. Imag.*, vol. 37, no. 6, pp. 1358–1369, Jun. 2018.
[159] J. Adler and O. Öktem, "Learned primal-dual reconstruction," *IEEE Trans. Med. Imag.*, vol. 37, no. 6, pp. 1322–1332, Jun. 2018.
[160] J. M. Wolterink, T. Leiner, M. A. Viergever, and

I. Išgum, "Generative adversarial networks for noise reduction in low-dose CT," *IEEE Trans. Med. Imag.*, vol. 36, no. 12, pp. 2536–2545, Dec. 2017.
[161] K. H. Jin, M. T. McCann, E. Froustey, and M. Unser, "Deep convolutional neural network for inverse problems in imaging," *IEEE Trans. Image Process.*, vol. 26, no. 9, pp. 4509–4522, Sep. 2017.
[162] Y. Han and J. C. Ye, "Framing U-net via deep convolutional framelets: Application to sparse-view CT," *IEEE Trans. Med. Imag.*, vol. 37, no. 6, pp. 1418–1429, Jun. 2018.
[163] S. Wang *et al.*, "Accelerating magnetic resonance imaging via deep learning," in *Proc. IEEE 13th Int. Symp. Biomed. Imag. (ISBI)*, Apr. 2016, pp. 514–517.
[164] K. Hammernik *et al.*, "Learning a variational network for reconstruction of accelerated MRI data," *Magn. Reson. Med.*, vol. 79, no. 6, pp. 3055–3071, 2017.
[165] Y. Han, J. Yoo, H. H. Kim, H. J. Shin, K. Sung, and J. C. Ye, "Deep learning with domain adaptation for accelerated projection-reconstruction MR," *Magn. Reson. Med*, vol. 80, no. 3, pp. 1189–1205, Sep. 2018. doi: 10.1002/mrm.27106.
[166] 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, p. 487, 2018.
[167] K. Gong *et al.*, "Iterative PET image reconstruction using convolutional neural network representation," *IEEE Trans. Med. Imag.*, vol. 38, no. 3, pp. 675–685, Mar. 2019.
[168] K. Gong, C. Catana, J. Qi, and Q. Li, "PET image reconstruction using deep image prior," *IEEE Trans. Med. Imag.*, vol. 38, no. 7, pp. 1655–1665, Jul. 2019.
[169] A. C. Luchies and B. C. Byram, "Deep neural networks for ultrasound beamforming," *IEEE Trans. Med. Imag.*, vol. 37, no. 9, pp. 2010–2021, Sep. 2018.
[170] Y. H. Yoon, S. Khan, J. Huh, and J. C. Ye, "Efficient B-mode ultrasound image reconstruction from sub-sampled RF data using deep learning," *IEEE Trans. Med. Imag.*, vol. 38, no. 2, pp. 325–336, Feb. 2019.
[171] Y. Rivenson, Z. Göröcs, H. Günaydin, Y. Zhang, H. Wang, and A. Ozcan, "Deep learning microscopy," *Optica*, vol. 4, no. 11, pp. 1437–1443, 2017.
[172] E. Nehme, L. E. Weiss, T. Michaeli, and Y. Shechtman, "Deep-STORM: Super-resolution single-molecule microscopy by deep learning," *Optica*, vol. 5, no. 4, pp. 458–464, 2018.
[173] A. Sinha, J. Lee, S. Li, and G. Barbastathis, "Lensless computational imaging through deep learning," *Optica*, vol. 4, no. 9, pp. 1117–1125, 2017.
[174] FDA. (2019). *510k Premarket Notification of AiCE Deep Learning Reconstruction (Canon)*. [Online]. Available: https://www.accessdata.fda.gov/scripts/cdrh/cfdocs/cfpmn/pmn.cfm?ID=K183046
[175] FDA. (2019). *510k Premarket Notification of Deep Learning Image Reconstruction (GE Medical Systems)*. [Online]. Available: https://www.accessdata.fda.gov/scripts/cdrh/cfdocs/cfpmn/pmn.cfm?ID=K183202
[176] C. H. McCollough *et al.*, "Low-dose CT for the detection and classification of metastatic liver lesions: Results of the 2016 Low Dose CT Grand Challenge," *Med. Phys.*, vol. 44, no. 10, pp. e339–e352, 2017.
[177] J. Zbontar *et al.*, "fastMRI: An open dataset and benchmarks for accelerated MRI," 2018, *arXiv:1811.08839*. [Online]. Available: https://arxiv.org/abs/1811.08839
[178] K. Kwon, D. Kim, and H. Park, "A parallel MR imaging method using multilayer perceptron," *Med. Phys.*, vol. 44, no. 12, pp. 6209–6224, 2017.
[179] H. Chen *et al.*, "Low-dose CT via convolutional neural network," *Biomed. Opt. Express*, vol. 8, no. 2, pp. 679–694, 2017.
[180] D. Lee, J. Yoo, and J. C. Ye, "Deep residual learning for compressed sensing MRI," in *Proc. IEEE 14th Int. Symp. Biomed. Imag. (ISBI)*, Apr. 2017, pp. 15–18.
[181] Q. Yang *et al.*, "Low-dose CT image denoising using a generative adversarial network with wasserstein distance and perceptual loss," *IEEE Trans. Med. Imag.*, vol. 37, no. 6, pp. 1348–1357, Jun. 2018.
[182] H. Gupta, K. H. Jin, H. Q. Nguyen, M. T. McCann, and M. Unser, "CNN-based projected gradient descent for consistent image reconstruction," 2017, *arXiv:1709.01809*. [Online]. Available: https://arxiv.org/abs/1709.01809
[183] T. M. Quan, T. Nguyen-Duc, and W.-K. Jeong, "Compressed sensing MRI reconstruction using a generative adversarial network with a cyclic loss," *IEEE Trans. Med. Imag.*, vol. 37, no. 6, pp. 1488–1497, Jun. 2018.
[184] H. Chen *et al.*, "LEARN: Learned experts' assessment-based reconstruction network for sparse-data CT," *IEEE Trans. Med. Imag.*, vol. 37, no. 6, pp. 1333–1347, Jun. 2018.
[185] H. K. Aggarwal, M. P. Mani, and M. Jacob, "MoDL: Model based deep learning architecture for inverse problems," 2017, *arXiv:1712.02862*. [Online]. Available: https://arxiv.org/abs/1712.02862
[186] D. Wu, K. Kim, G. El Fakhri, and Q. Li, "Iterative low-dose CT reconstruction with priors trained by artificial neural network," *IEEE Trans. Med. Imag.*, vol. 36, no. 12, pp. 2479–2486, Dec. 2017.
[187] K. Gong *et al.*, "Iterative PET image reconstruction using convolutional neural network representation," 2017, *arXiv:1710.03344*. [Online]. Available: https://arxiv.org/abs/1710.03344
[188] K. Gregor and Y. LeCun, "Learning fast approximations of sparse coding," in *Proc. 27th Int. Conf. Int. Conf. Mach. Learn.*, 2010, pp. 399–406.
[189] A. Beck and M. Teboulle, "A fast iterative shrinkage-thresholding algorithm for linear inverse problems," *SIAM J. Imag. Sci.*, vol. 2, no. 1, pp. 183–202, 2009.
[190] D. Ulyanov, A. Vedaldi, and V. Lempitsky, "Deep image prior," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2018, pp. 9446–9454.
[191] T. Würfl, F. C. Ghesu, V. Christlein, and A. Maier, "Deep learning computed tomography," in *Proc. Int. Conf. Med. Image Comput. Comput.-Assist. Intervent*. Cham, Switzerland: Springer, 2016, pp. 432–440.
[192] M. U. Ghani and W. C. Karl, "Fast enhanced CT metal artifact reduction using data domain deep learning," 2019, *arXiv:1904.04691*. [Online]. Available: https://arxiv.org/abs/1904.04691
[193] M. U. Ghani and W. C. Karl, "Deep learning based sinogram correction for metal artifact reduction," *Electron. Imag.*, vol. 2018, no. 15, pp. 472-1–472-8, 2018.
[194] H. Lee, J. Lee, H. Kim, B. Cho, and S. Cho, "Deep-neural-network-based sinogram synthesis for sparse-view CT image reconstruction," *IEEE Trans. Radiat. Plasma Med. Sci.*, vol. 3, no. 2, pp. 109–119, Mar. 2019.
[195] M. U. Ghani and W. C. Karl, "Deep learning-based sinogram completion for low-dose CT," in *Proc. IEEE 13th Image, Video, Multidimensional Signal Process. Workshop (IVMSP)*, Jun. 2018, pp. 1–5.
[196] Y. Han, L. Sunwoo, and J. C. Ye, "k-space deep learning for accelerated MRI," *IEEE Trans. Med. Imag.*, to be published. [Online]. Available: https://arxiv.org/abs/1805.03779
[197] J. Lee, Y. Han, J.-K. Ryu, J.-Y. Park, and J. C. Ye, "k-space deep learning for reference-free EPI ghost correction," *Magn. Reson. Med.*, to be published. doi: 10.1002/mrm.27896.
[198] M. Akçakaya, S. Moeller, S. Weingärtner, and K. Uğurbil, "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, 2019.
[199] Y. Han and J. C. Ye, "One network to solve all ROIs: Deep learning CT for any ROI using differentiated backprojection," *Med. Phys.*, to be published. [Online]. Available: https://arxiv.org/abs/1810.00500
[200] E. Kang, H. J. Koo, D. H. Yang, J. B. Seo, and J. C. Ye, "Cycle-consistent adversarial denoising network for multiphase coronary CT angiography," *Med. Phys.*, vol. 46, no. 2, pp. 550–562, Feb. 2018.
[201] R. J. Duffin and A. C. Schaeffer, "A class of nonharmonic Fourier series," *Trans. Amer. Math. Soc.*, vol. 72, no. 2, pp. 341–366, Mar. 1952.
[202] R. Yin, T. Gao, Y. M. Lu, and I. Daubechies, "A tale of two bases: Local-nonlocal regularization on image patches with convolution framelets," *SIAM J. Imag. Sci.*, vol. 10, no. 2, pp. 711–750, 2017.
[203] E. J. Candès, L. Demanet, D. Donoho, and L. Ying, "Fast discrete curvelet transforms," *Multiscale Model. Simul.*, vol. 5, no. 3, pp. 861–899, Sep. 2006.
[204] M. N. Do and M. Vetterli, "The contourlet transform: An efficient directional multiresolution image representation," *IEEE Trans. Image Process.*, vol. 14, no. 12, pp. 2091–2106, Dec. 2005.
[205] J. C. Ye, Y. Han, and E. Cha, "Deep convolutional framelets: A general deep learning framework for inverse problems," *SIAM J. Imag. Sci.*, vol. 11, pp. 991–1048, Jan. 2018.
[206] J. C. Ye and W. K. Sung, "Understanding geometry of encoder-decoder CNNs," in *Proc. 36th Int. Conf. Mach. Learn. (PMLR)*, Long Beach, CA, USA, vol. 97, K. Chaudhuri and R. Salakhutdinov, Eds., Jun. 2019, pp. 7064–7073.
[207] H. H. Barrett, C. K. Abbey, and E. Clarkson, "Objective assessment of image quality. III. ROC metrics, ideal observers, and likelihood-generating functions," *J. Opt. Soc. Amer. A, Opt. Image Sci.*, vol. 15, no. 6, pp. 1520–1535, Jun. 1998.
[208] Z. Wang and A. C. Bovik, "Mean squared error: Love it or leave it? A new look at signal fidelity measures," *IEEE Signal Process. Mag.*, vol. 26, no. 1, pp. 98–117, Jan. 2009.
[209] S. Ye, S. Ravishankar, Y. Long, and J. A. Fessler, "SPULTRA: Low-dose CT image reconstruction with joint statistical and learned image models," 2018, *arXiv:1808.08791*. [Online]. Available: https://arxiv.org/abs/1808.08791
[210] S. Ravishankar and Y. Bresler, "Adaptive sampling design for compressed sensing MRI," in *Proc. Annu. Int. Conf. IEEE Eng. Med. Biol. Soc.*, Aug./Sep. 2011, pp. 3751–3755.
[211] B. Gözcü *et al.*, "Learning-based compressive MRI," *IEEE Trans. Med. Imag.*, vol. 37, no. 6, pp. 1394–1406, Jun. 2018.
[212] Q. Nguyen and M. Hein, "Optimization landscape and expressivity of deep CNNs," in *Proc. Int. Conf. Mach. Learn.*, 2018, pp. 3727–3736.
[213] R. Ge and T. Ma, "On the optimization landscape of tensor decompositions," in *Proc. Adv. Neural Inf. Process. Syst.*, 2017, pp. 3653–3663.
[214] S. S. Du, J. D. Lee, Y. Tian, B. Poczos, and A. Singh, "Gradient descent learns one-hidden-layer CNN: Don't be afraid of spurious local minima," in *Proc. Int. Conf. Mach. Learn.*, 2018, pp. 1338–1347.
[215] S. Gunasekar, J. D. Lee, D. Soudry, and N. Srebro, "Implicit bias of gradient descent on linear convolutional networks," in *Proc. Adv. Neural Inf. Process. Syst.*, 2018, pp. 9461–9471.
[216] D. Soudry, E. Hoffer, M. S. Nacson, S. Gunasekar, and N. Srebro, "The implicit bias of gradient descent on separable data," *J. Mach. Learn. Res.*, vol. 19, no. 1, pp. 2822–2878, 2018.

ABOUT THE AUTHORS

Saiprasad Ravishankar (Member, IEEE) received the B.Tech. degree in electrical engineering from IIT Madras, Chennai, India, in 2008, and the M.S. and Ph.D. degrees in electrical and computer engineering from the University of Illinois at Urbana–Champaign, Champaign, IL, USA, in 2010 and 2014, respectively.

He was then an Adjunct Lecturer with the Department of Electrical and Computer Engineering, University of Illinois at Urbana-Champaign, and a Postdoctoral Research Associate with the Coordinated Science Laboratory. Since 2015, he was a Research Fellow with the Department of Electrical Engineering and Computer Science, University of Michigan, Ann Arbor, MI, USA. From August 2018 to February 2019, he was a Postdoctoral Research Associate with the Theoretical Division, Los Alamos National Laboratory, Los Alamos, NM, USA. He is currently an Assistant Professor with the Departments of Computational Mathematics, Science and Engineering, and Biomedical Engineering, Michigan State University, East Lansing, MI, USA. His current research interests include signal and image processing, biomedical and computational imaging, data-driven methods, machine learning, signal modeling, inverse problems, data science, compressed sensing, and large-scale data processing.

Dr. Ravishankar is currently a member of the IEEE Computational Imaging Technical Committee. He was a recipient of the IEEE Signal Processing Society Young Author Best Paper Award in 2016. His coauthored paper received the Best Student Paper Award at the IEEE International Symposium on Biomedical Imaging (ISBI), in 2018, and another was a Finalist at the IEEE International Workshop on Machine Learning for Signal Processing (MLSP), in 2017.

Jong Chul Ye (Senior Member, IEEE) received the B.Sc. and M.Sc. degrees from Seoul National University, Seoul, South Korea, in 1993 and 1995, respectively, and the Ph.D. from Purdue University, West Lafayette, IN, USA, in 1999.

From 1999 to 2001, he was a Postdoctoral Fellow with the Coordinated Science Laboratory, University of Illinois at Urbana–Champaign, Champaign, IL, USA. From 2001 to 2004, he was with Philips Research, Briarcliff Manor, NY, USA, and GE Global Research, Niskayuna, NY, USA. In 2004, he joined Korea Advanced Institute of Science and Technology (KAIST), Daejeon, South Korea, as an Assistant Professor, where he is currently a Professor with the Department of Bio/Brain Engineering and also an Adjunct Professor with the Department of Mathematical Sciences. His current research interests include machine learning and sparse recovery for various imaging reconstruction problems in X-ray CT, MRI, optics, ultrasound, and so on.

Dr. Ye's group was the first place winner of the 2009 Recon Challenge at the ISMRM Workshop, and the second place winner at the 2016 Low Dose CT Grand Challenge organized by the American Association of Physicists in Medicine (AAPM). He was an Advisor of Best Student Paper Awards (first and runner-up) at the 2013 and 2016 IEEE Symposium on Biomedical Imaging (ISBI). He is currently the Vice Chair of the IEEE SPS Technical Committee for Computational Imaging, and the General Co-Chair (with Mathews Jacob) of the 2020 ISBI, Iowa City. He has served as an Associate Editor for the IEEE TRANSACTIONS ON IMAGE PROCESSING, the IEEE TRANSACTIONS ON COMPUTATIONAL IMAGING, the *Journal of Electronic Imaging*, an Editorial Board Member for *Magnetic Resonance in Medicine*, and an International Advisory Board Member for *Physics in Medicine and Biology*. He is currently an Associate Editor of the IEEE TRANSACTIONS ON MEDICAL IMAGING, a Senior Editor of the *IEEE Signal Processing Magazine*, and a Section Editor of *BMC Biomedical Engineering*.

Jeffrey A. Fessler (Fellow, IEEE) received the B.S.E.E. degree from Purdue University, West Lafayette, IN, USA, in 1985, and the M.S.E.E. degree, the M.S. degree in statistics, and the Ph.D. degree in electrical engineering from Stanford University, Stanford, CA, USA, in 1986, 1989, and 1990, respectively.

He has supervised doctoral research in PET, SPECT, X-ray CT, MRI, and optical imaging problems. From 1985 to 1988, he was a National Science Foundation Graduate Fellow with Stanford University. Since 1990, he has been with the University of Michigan, Ann Arbor, MI, USA. From 1991 to 1992, he was a Department of Energy Alexander Hollaender Post-Doctoral Fellow with the Division of Nuclear Medicine. From 1993 to 1995, he was an Assistant Professor of the Nuclear Medicine and the Bioengineering Program. He is currently a Professor with the Departments of Electrical Engineering and Computer Science, Radiology, and Biomedical Engineering, where he is also the William L. Root Professor with the Department of Electrical Engineering and Computer Sciences (EECS). His current research interests include statistical aspects of imaging problems.

Dr. Fessler became a fellow of the IEEE in 2006, for contributions to the theory and practice of image reconstruction. He was a recipient of the Francois Erbsmann Award for his IPMI93 presentation and the Edward Hoffman Medical Imaging Scientist Award in 2013. He was the Chair of the IEEE T-MI Steering Committee and the ISBI Steering Committee. He was the Co-Chair of the 1997 SPIE Conference on Image Reconstruction and Restoration, the Technical Program Co-Chair of the 2002 IEEE International Symposium on Biomedical Imaging (ISBI), and the General Chair of ISBI 2007. He has served as an Associate Editor for the IEEE TRANSACTIONS ON MEDICAL IMAGING, IEEE SIGNAL PROCESSING LETTERS, the IEEE TRANSACTIONS ON IMAGE PROCESSING, and the IEEE TRANSACTIONS ON COMPUTATIONAL IMAGING. He is currently serving as an Associate Editor for the *SIAM Journal on Imaging Science*.