2018 IEEE 15th International Symposium on Biomedical Imaging (ISBI 2018)
April 4-7, 2018, Washington, D.C., USA

DEEP DICTIONARY-TRANSFORM LEARNING FOR IMAGE RECONSTRUCTION

Saiprasad Ravishankar   Anish Lahiri   Cameron Blocker   Jeffrey A. Fessler

Department of Electrical Engineering and Computer Science, University of Michigan, Ann Arbor, MI, USA

ABSTRACT

Various dictionaries and transforms are known to sparsify images, and have been exploited in algorithms for image reconstruction from limited data. Methods that simultaneously reconstruct the image and learn dictionaries or sparsifying transforms for image patches, called blind compressed sensing methods, have shown promising performance. Motivated by such adaptive algorithms, this paper proposes an approach to train dictionary-transform based methods for image reconstruction by minimizing a minimum absolute reconstruction error criterion. Each "layer" of the algorithm consists of applying (convolving) trained transforms, thresholdings, and dictionaries to images, followed by a simple least squares update of the images. Numerical experiments illustrate the usefulness and speed-ups provided by such trained algorithms compared to related schemes.

***Index Terms*—** Sparsifying transforms, Dictionaries, Machine learning, Sparse representations, Smart imaging.

1. INTRODUCTION

Imaging approaches such as magnetic resonance imaging (MRI) and X-ray computed tomography (CT) often involve image reconstruction from limited or corrupted data such as in the case of low-dose CT, sparse-view CT, compressive sensing [1, 2] based static MRI, or dynamic MRI where the data is inherently undersampled. Imaging with limited data often has potential benefits such as scan time speed-ups for MRI or reduced radiation exposure for X-ray CT.

Several image reconstruction methods have been proposed [3–7] exploiting image characteristics such as the sparsity of images in analytical dictionaries or transform domains, low-rank properties [8, 9], etc. More recently, data-driven approaches for image reconstruction that learn synthesis dictionaries or sparsifying transforms have received attention [10–14] and have shown promise in medical imaging applications. Sparsifying transform-based [15] approaches are typically faster than dictionary-based approaches due to relatively inexpensive closed-form thresholding-based sparse coding, whereas synthesis sparse coding is NP-hard in general. The transforms or dictionaries could be learned from training data [14] (using various transform learning [15–17] or dictionary learning [18–20] methods) and used to reconstruct other images, or even learned simultaneously while reconstructing the images [12], which is called (transform or dictionary) blind compressed sensing.

Motivated by the efficient (convolutional) structure of recent transform-blind compressed sensing algorithms [12, 13], this paper proposes an approach for training the parameters (e.g., filters, thresholds, etc.) of transform-based algorithms for image reconstruction, by minimizing a minimum absolute reconstruction error cost.[1] Each iteration (called layer) of the algorithm consists of applying trained complex-valued transforms, thresholdings (non-linearities), and dictionaries to images, followed by a simple (full) least squares update of the images that takes into account the imaging model or physics. Numerical results illustrate potential for the trained method over related schemes for image reconstruction. Very recent works [22,23] also proposed incorporating the imaging model in training specific algorithms for reconstruction. However, our proposed method differs in the specific algorithm architecture/updates, non-linearities, the training objective, and algorithm for training.

2. IMAGE RECONSTRUCTION MODEL AND TRAINING ALGORITHM

This section describes the proposed reconstruction algorithm architecture and its training. We first briefly describe a recent transform-blind compressed sensing approach (UTMRI) [13] that motivates our proposed method.

2.1. Background and Reconstruction Model

The goal in inverse problems is to estimate an unknown signal or (vectorized) image $x \in \mathbb{C}^p$ from its (typically limited or corrupted) measurements $y \in \mathbb{C}^m$. A typical regularized inverse problem is as follows:

$$\underset{x \in \mathbb{C}^p}{\arg\min} \; \|Ax - y\|_2^2 + \zeta(x) \tag{1}$$

This work was supported in part by the following grants: ONR grant N00014-15-1-2141, DARPA Young Faculty Award D14AP00086, ARO MURI grants W911NF-11-1-0391 and 2015-05174-05, NIH grant R01 EB023618, and a UM-SJTU seed grant.

[1] A recent work [21] has shown promise for minimizing MAE (ℓ_1) costs over mean squared error or MSE (ℓ_2) costs for model training. The former may enable better generalization of learned models.

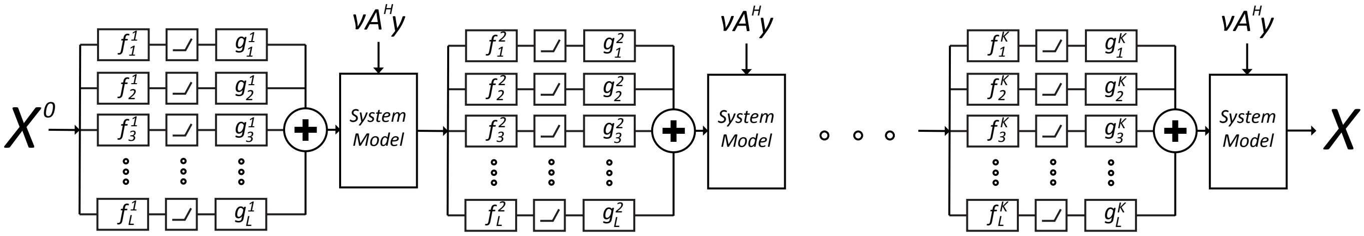

Fig. 1. Proposed image reconstruction architecture based on the image update step of UTMRI [13]. The initial image (x^0) is iteratively passed through a decorruption step involving filtering and thresholding operations, followed by least squares image update (system model block) that enforces the imaging forward model. The decorruption step involves a bank of transform filters (f_l^k, with k denoting iteration and l the filter number), thresholdings, and dictionary filters (g_l^k). For UTMRI, both the transform and dictionary filters are related to the rows of W (matched) and $L = n$.

where $A \in \mathbb{C}^{m\times p}$ is the imaging measurement matrix (e.g., a Fourier encoding for MRI) and $\zeta(x)$ is a regularizer capturing assumed properties of the image (e.g., sparsity in a transform domain). In the blind compressed sensing framework, a data-driven (e.g., involving transform learning) regularizer is used enabling joint image reconstruction and image model learning. Here, we draw inspiration from the recent unitary transform-blind compressed sensing scheme involving the following optimization problem [13]:

$$\text{(P0)} \min_{x,W,B} \sum_{j=1}^{N} \|WP_jx - b_j\|_2^2 + \nu\|Ax-y\|_2^2 + \gamma^2\|B\|_0, \quad \text{s.t. } W^HW = I,$$

where $\nu, \gamma > 0$ are parameters, P_j is an operator that extracts the jth patch (N overlapping patches assumed) of image x as a vector $P_jx \in \mathbb{C}^n$, and $W \in \mathbb{C}^{n\times n}$ is a sparsifying transform such that $WP_jx \approx b_j$ with sparse b_j. Matrix $B \in \mathbb{C}^{n\times N}$ has the sparse b_j's as its columns, and the ℓ_0 "norm" counts the number of non-zeros in a vector or matrix. Operations $(\cdot)^H$ and $(\cdot)^T$ denote conjugate transpose and the usual transpose, respectively. Problem (P0) (cf. [12] for variants) allows the transform to adapt to the underlying or imaged object.

Prior work [12,13] proposed an iterative block coordinate descent (BCD) algorithm for (P0) that alternates between solving for W (*transform update step*), B (*sparse coding step*), and x (*image update step*) with other variables fixed. Here, the sparse codes are updated as $\hat{b}_j = H_\gamma(WP_jx)$, where $H_\gamma(\cdot)$ is the hard-thresholding operator that sets vector or matrix entries with magnitude less than γ to zero while leaving other entries unaffected. The output in the transform update step is $\hat{W} = VU^H$, where $PB^H = U\Sigma V^H$ denotes a full singular value decomposition (SVD) with P being the matrix whose columns are the vectorized (and most recently estimated) image patches. The image update step for (P0) involves a least squares problem whose normal equation is as follows:

$$Gx^k = \nu A^H y + \sum_{j=1}^{N} P_j^T D^k H_\gamma(W^k P_j x^{k-1}) \qquad (2)$$

where k denotes the BCD iteration number, $D^k \triangleq \left(W^k\right)^H$, and $G \triangleq \sum_{j=1}^N P_j^T P_j + \nu A^H A$ is a fixed matrix. For single-coil Cartesian MRI, where $A = F_u$, the undersampled Fourier encoding, G is readily diagonalized by the 2D DFT and the update in (2) is performed cheaply using FFTs [13]. Alternatively, one could solve (2) using iterative methods such as conjugate gradients (CG).

Fig. 1 provides a schematic of the image update over the BCD iterations for (P0). Note that the second term in (2) is equivalently written as $\sum_{l=1}^n \sum_{j=1}^N P_j^T d_l^k H_\gamma(r_l^{k^T} P_j x^{k-1})$, with d_l and r_l denoting the lth columns of D and $R = W^T$, respectively. In particular, when all the overlapping image patches (with patch stride of 1 pixel) are used in (P0) and the patches overlapping the image boundaries wrap around on the opposite side of the image (periodic image condition), then the second term in (2) involves filtering (via circular convolution) x^{k-1} with each transform filter (the filter is obtained by flipping and zero-padding a reshaped row of W), hard-thresholding the result, filtering with each corresponding dictionary (D) filter (obtained by zero-padding the reshaped column of D), and aggregating the outputs from the various ($1 \le l \le n$) filters. This aggregate is further updated by adding the bias term $\nu A^H y$ and applying G^{-1} (or alternatively performing CG).

The filtering and thresholding operations (with the filters learned via the aforementioned SVD-based update using (P0)) in each iteration (or layer) of Fig. 1 help denoise or decorrupt the image and the other operations enforce the imaging model or physics. The (UTMRI) algorithm thus has a convolutional network architecture (with depth corresponding to the number of iterations), where the filters are *learned* *on-the-fly* from the measurements y using (P0) (i.e., without using any explicit training data). Next, we will describe an alternative approach to train such an algorithm using training datasets to explicitly minimize image reconstruction errors. Such pre-training would save runtime when applying the algorithm to new test data.

2.2. Training Cost and Algorithm

Our proposed method exploits the iterative or multi-layer structure in Fig. 1, with each layer comprised of a *trained* decorruption step, followed by the least squares update in (2) that uses the imaging system model. Learning the decorruption step involves learning the transform and dictionary filters and thresholdings. We assume L transform and dictionary filters in each layer, and a potentially different threshold for each filter pair (or arm in Fig. 1). We use soft-thresholding, which is amenable to gradient-based optimization. A complex-valued scalar is soft-thresholded as $S_\tau(c) = \max(|c| - \tau, 0)e^{j\angle c}$ for $\tau \geq 0$.

The data for training consists of a set of reference (ground truth) images reconstructed from densely sampled imaging measurements along with subsampled measurements of these images. Initial images are obtained from the measurements (e.g., $A^\dagger y$), and the dictionary, transform, and thresholds are trained layer-by-layer. Once the parameters for a specific layer have been trained, the training image reconstructions are passed through that layer (with fixed parameters) and the resulting images are used for training the next layer. In the kth layer, we minimize the following cost consisting of an ℓ_1 (with $\|Q\|_1 = \|\text{vec}(Q)\|_1$) patch-based reconstruction error and a regularizer (a related cost appears in [24]), to train the model:

$$\text{(P1)} \quad \min_{D^k, W^k, \Gamma^k} \|P^{\text{train}} - D^k S_{\Gamma^k}(W^k \hat{P}^{k-1})\|_1 + \beta \tilde{R}(D^k),$$

where $D^k \in \mathbb{C}^{n\times L}$, $W^k \in \mathbb{C}^{L\times n}$ and $\Gamma^k \in \mathbb{R}^L$ are the dictionary, transform (with L filters), and a set of thresholds, respectively, for the kth reconstruction layer. The regularizer $\tilde{R}(D^k) \triangleq \sum_{l=1}^{L} \left(\|d_l^k\|_2^2 - 1\right)^2$ with non-negative weight β keeps the ℓ_2 norms of dictionary columns (d_l^k) close to unity, which helps eliminate scaling ambiguities[2] [25] in the solution. The operator S_{Γ^k} performs entry-wise soft-thresholding using corresponding thresholds. Matrix $P^{\text{train}} \in \mathbb{C}^{n\times J}$ has a set of (often randomly chosen) patches from reference (ground truth) images as its columns, and $\hat{P}^{k-1}$ has the corresponding patches from the reconstructed versions of the images after layer $k-1$.

[2] Each column of D^k can be scaled by a positive constant α and the corresponding row of W^k and entry of Γ^k can be both scaled by $1/\alpha$ (the soft-thresholding function is scale-homogenous jointly with respect to the row of W^k and corresponding entry of Γ^k), and the cost in (P1) is invariant to such scalings.

Unlike the previous UTMRI algorithm, the proposed architecture uses soft-thresholding instead of hard-thresholding. We do not however constrain D^k and W^k to be "matched", which allows a degree of flexibility for the proposed scheme.

The variables D^k, W^k, and Γ^k could be updated using gradient-based techniques such as sub-derivative descent with backtracking line search for step sizes. This would ensure monotone decrease of the cost. Here, we instead use the recent ADAM approach [26] that exploits higher-order gradient information to jointly update the variables. ADAM performs stochastic updates, where each iteration computes the gradient of the objective corresponding to a subset (minibatch) of all data (thus saving memory). The following are gradients (computed efficiently using sparse multiplications) of the cost in (P1) with respect to the columns d_l^k and r_l^k, and scalar γ_l^k (i.e., lth entry of Γ^k), where $R = W^T$:

$$\frac{\partial \psi}{\partial \gamma_l^k} = \text{Re}\left(d_l^{k^T} \text{sign}^*(E^k)\, h_l^{k^T}\right)$$

$$\frac{\partial \psi}{\partial d_l^k} = -\text{sign}\left(E^k\right) S_{\gamma_l^k}^H(c_l^k) + 4\beta\left(\|d_l^k\|_2^2 - 1\right) d_l^k$$

$$\frac{\partial \psi}{\partial r_l^{k^T}} = \left(\phi - \gamma_l^k \phi \oslash |c_l^k| + \gamma_l^k \text{Re}\big(\phi \odot \text{sign}^*(c_l^k)\big) \oslash c_l^k\right) \times \left(\hat{Z} \odot \mathbf{1}\, \mathbf{1}_{|c_l^k| > \gamma_l^k}\right)^H$$

where $E^k \triangleq P^{\text{train}} - D^k S_{\Gamma^k}(R^{k^T} \hat{P}^{k-1})$, $c_l^k \triangleq r_l^{k^T} \hat{P}^{k-1}$, $h_l^k \triangleq \text{sign}(c_l^k) \odot \mathbf{1}_{|c_l^k| > \gamma_l^k}$, $\text{sign}(\cdot)$ denotes the complex phase computed element-wise, $\mathbf{1}$ and $\mathbf{1}_{|c_l^k| > \gamma_l^k}$ denote a length-n column vector of ones and the (row vector) indicator function (takes value 0 when condition is violated and 1 otherwise) computed element-wise, $\phi \triangleq (d_l^k)^H \text{sign}(E^k)$ and ψ is the cost in (P1). Here, $\odot$ denotes element-wise multiplication, $\oslash$ denotes element-wise division, $(\cdot)^*$ denotes the complex conjugate, and $\text{Re}(\cdot)$ denotes real part.

Once all layers of the algorithm have been trained by optimizing (P1) for each successive k, we reconstruct new test data by passing the initial image reconstruction from undersampled measurements through the trained network.

3. NUMERICAL EXPERIMENTS

Here, we present preliminary experiments illustrating the performance of the proposed method for MR image reconstruction. We used the multi-slice dataset with 32 512×512 (complex-valued) slices provided by Prof. Michael Lustig, UC Berkeley. We simulated single coil Cartesian k-space measurements from a subset (five) of these slices, and used these to train our proposed image reconstruction model with 20 layers at 3.3 fold and 5 fold undersampling of k-space, and $\beta = 10^4$. The transforms and dictionaries in each layer had 256 filters, which were trained together with the corresponding 256 soft-thresholds, from 8×8 patches, and $\nu = 10^6/p$ with p the number of image pixels [13]. We used variable

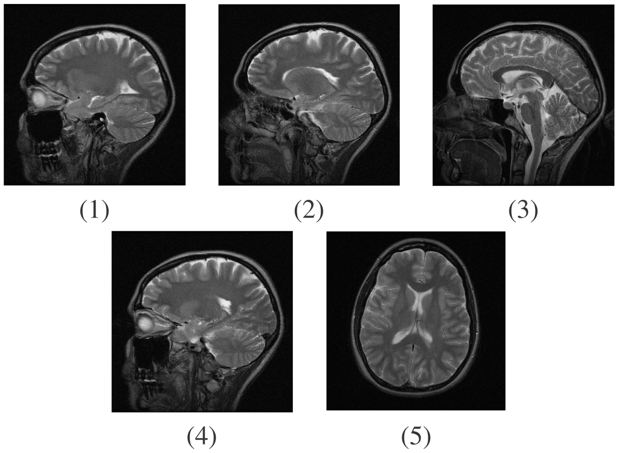

Fig. 2. 512×512 test images used in Table 1.

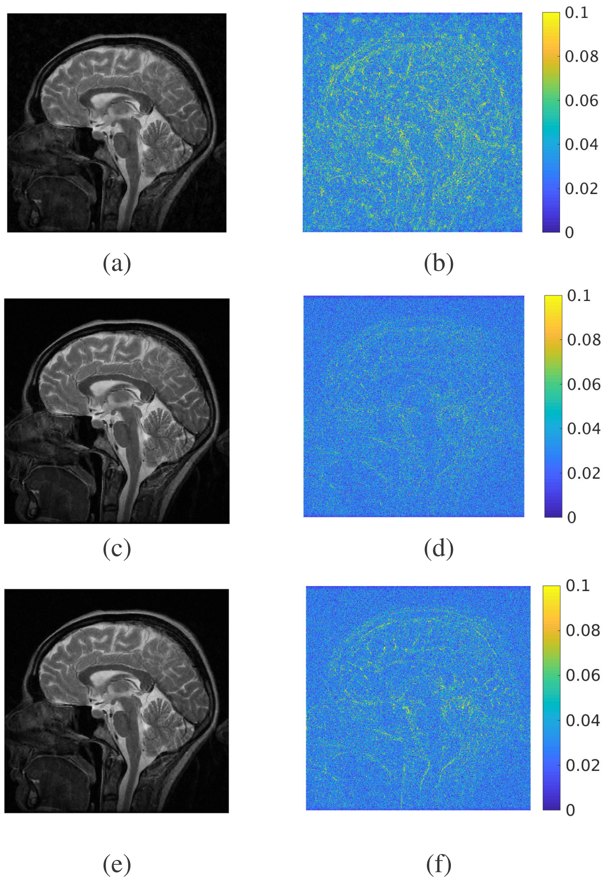

Fig. 3. Reconstructed images (magnitudes shown) and reconstruction error maps (magnitude of difference between reconstructed and reference images) for image 3 with 30% samples using Sparse MRI (top row), UTMRI (middle row) and our trained method (bottom row).

UF	Image	Zero-filled	Sparse MRI	UTMRI	Proposed
3.3x	1	25.6	26.7	28.3	28.2
	2	25.2	26.6	27.9	27.8
	3	26.0	27.3	29.3	28.9
	4	25.4	26.7	28.2	28.1
	5	27.2	28.9	30.6	30.3
5x	1	24.7	25.9	27.6	27.5
	2	24.2	25.5	27.2	27.0
	3	24.9	26.3	28.5	28.0
	4	24.4	25.7	27.6	27.4
	5	26.2	27.9	29.8	29.5

Table 1. PSNR values in decibels (dB) for the five test images at two undersampling factors (UF) using various methods.

density 2D random sampling patterns [27], which are feasible for 3D imaging.

The trained models were used to reconstruct five test images from undersampled measurements, including an axial slice from a different acquisition (Fig. 2). We compare the performance of the trained method to Sparse MRI [3] that exploits sparsity in wavelets and total variation domains, and the recent blind compressed sensing method UTMRI [13]. We ran UTMRI for 120 iterations (a transform is estimated in each iteration), with other parameters as in prior work [13]. We used the built-in parameter settings in the publicly available Sparse MRI implementation [28]. These settings performed well in our experiments. We evaluated reconstruction quality on test data using the peak signal to noise ratio (PSNR) metric (in decibels (dB)) computed between the complex-valued reference and reconstructed images.

Table 1 lists the PSNR values for zero-filled IFFT (the initial image for the methods), Sparse MRI, UTMRI, and the trained algorithm for various test images and undersampling factors. Both UTMRI and the trained method outperformed the non-adaptive Sparse MRI. The trained method performed quite similarly as UTMRI but was up to 5x faster (average reconstruction times for UTMRI and Sparse MRI were about 241s and 100s respectively, compared to about 50s for our method). While our current implementation uses patch-based processing for testing, much lower runtimes can be achieved using fast convolutional implementations. Figure 3 compares the reconstructed images and error maps for different methods showing potential for our trained approach.

4. CONCLUSIONS

In this work, motivated by recent works on transform-blind compressed sensing, we presented a deep dictionary-transform learning method for image reconstruction using a minimum absolute error criteron. The image reconstruction algorithm learned from training scans can be applied relatively cheaply to reconstruct other images from undersampled measurements. Preliminary experiments illustrated potential for such a trained algorithm in terms of image quality and runtime. Investigation of end-to-end training (instead of layer-by-layer) of the reconstruction scheme, and more extensive evaluation and validation on large datasets will be performed in future work.

5. REFERENCES

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

[2] D. Donoho, "Compressed sensing," *IEEE Trans. Information Theory*, vol. 52, no. 4, pp. 1289–1306, 2006.

[3] M. Lustig, D.L. Donoho, and J.M. Pauly, "Sparse MRI: The application of compressed sensing for rapid MR imaging," *Magnetic Resonance in Medicine*, vol. 58, no. 6, pp. 1182–1195, 2007.

[4] M. Lustig, J. M. Santos, D. L. Donoho, and J. M. Pauly, "k-t SPARSE: High frame rate dynamic MRI exploiting spatio-temporal sparsity," in *Proc. ISMRM*, 2006, p. 2420.

[5] J. Trzasko and A. Manduca, "Highly undersampled magnetic resonance image reconstruction via homotopic l_0-minimization," *IEEE Trans. Med. Imaging*, vol. 28, no. 1, pp. 106–121, 2009.

[6] Y. Kim, M. S. Nadar, and A. Bilgin, "Wavelet-based compressed sensing using gaussian scale mixtures," in *Proc. ISMRM*, 2010, p. 4856.

[7] H. Jung, K. Sung, K. S. Nayak, E. Y. Kim, , and J. C. Ye, "k-t FOCUSS: A general compressed sensing framework for high resolution dynamic MRI," *Magnetic Resonance in Medicine*, vol. 61, no. 1, pp. 103–116, 2009.

[8] Z. P. Liang, "Spatiotemporal imaging with partially separable functions," in *IEEE Intl. Symp. Biomed. Im. (ISBI)*, 2007, pp. 988–991.

[9] J. P. Haldar and Z. P. Liang, "Spatiotemporal imaging with partially separable functions: A matrix recovery approach," in *IEEE Intl. Symp. Biomed. Im. (ISBI)*, 2010, pp. 716–719.

[10] 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, 2011.

[11] 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, Sept. 2012.

[12] S. Ravishankar and Y. Bresler, "Efficient blind compressed sensing using sparsifying transforms with convergence guarantees and application to magnetic resonance imaging," *SIAM Journal on Imaging Sciences*, vol. 8, no. 4, pp. 2519–2557, 2015.

[13] S. Ravishankar and Y. Bresler, "Data-driven learning of a union of sparsifying transforms model for blind compressed sensing," *IEEE Transactions on Computational Imaging*, vol. 2, no. 3, pp. 294–309, 2016.

[14] X. Zheng, S. Ravishankar, Y. Long, and J. A. Fessier, "Union of learned sparsifying transforms based low-dose 3D CT image reconstruction," in *International Conference on Fully 3D Image Reconstruction in Radiology and Nuclear Medicine*, 2017, pp. 69–72.

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

[16] S. Ravishankar, B. Wen, and Y. Bresler, "Online sparsifying transform learning – part I: Algorithms," *IEEE Journal of Selected Topics in Signal Processing*, vol. 9, no. 4, pp. 625–636, 2015.

[17] S. Ravishankar and Y. Bresler, "Learning doubly sparse transforms for images," *IEEE Trans. Image Process.*, vol. 22, no. 12, pp. 4598–4612, 2013.

[18] M. Aharon, M. Elad, and A. Bruckstein, "K-SVD: An algorithm for designing overcomplete dictionaries for sparse representation," *IEEE Transactions on signal processing*, vol. 54, no. 11, pp. 4311–4322, 2006.

[19] M. Yaghoobi, T. Blumensath, and M. Davies, "Dictionary learning for sparse approximations with the majorization method," *IEEE Transaction on Signal Processing*, vol. 57, no. 6, pp. 2178–2191, 2009.

[20] 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, 2010.

[21] H. Zhao, O. Gallo, I. Frosio, and J. Kautz, "Loss functions for image restoration with neural networks," *IEEE Transactions on Computational Imaging*, vol. 3, no. 1, pp. 47–57, Mar. 2017.

[22] J. Schlemper, J. Caballero, J. V. Hajnal, A. Price, and D. Rueckert, "A deep cascade of convolutional neural networks for dynamic mr image reconstruction," *IEEE Transactions on Medical Imaging*, 2018, to appear.

[23] K. Hammernik, T. Klatzer, E. Kobler, M. P. Recht, D. K. Sodickson, T. Pock, and F. Knoll, "Learning a variational network for reconstruction of accelerated MRI data," *Magnetic Resonance in Medicine*, 2018, to appear.

[24] R. Rubinstein and M. Elad, "Dictionary learning for analysis-synthesis thresholding," *IEEE Transactions on Signal Processing*, vol. 62, no. 22, pp. 5962–5972, 2014.

[25] R. Gribonval and K. Schnass, "Dictionary identification–sparse matrix-factorization via l_1 -minimization," *IEEE Trans. Inform. Theory*, vol. 56, no. 7, pp. 3523–3539, 2010.

[26] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," in *Proc. 3rd International Conference for Learning Representations*, 2015.

[27] B. Roman, A. Hansen, and B. Adcock, "On asymptotic structure in compressed sensing," *arXiv preprint arXiv:1406.4178*, 2014.

[28] M. Lustig, "Michael Lustig home page," `http://www.eecs.berkeley.edu/~mlustig/Software.html`, 2014, [Online; accessed October, 2017].