SYNTHESIS-ANALYSIS DECONVOLUTIONAL NETWORK FOR COMPRESSED SENSING

Qiegen Liu[1,2], Henry Leung[2]
[1]School of Electronic Information Engineering, Nanchang University, Nanchang, 330031, China
[2]Department of Electrical and Computer Engineering, University of Calgary, Calgary, T2N 1N4, Canada
E-mail: liuqiegen@ncu.edu.cn

ABSTRACT

Synthesis learning and analysis learning, with sparse coding (SC) and Markov random fields (MRFs) as two representative types of models, are two complementary tools to describe the image manifolds. SC has strengths in representing the regular features/explicit visual manifolds while its effectiveness depends on the training dataset. While MRFs have great potentials to characterize the stochastic textures/implicit visual manifolds but at the cost of high training complexity. In this paper, by means of the convolutional operator, a unified synthesis and analysis deconvolutional network (SADN) is presented. It not only requires the generative coding coefficients to be sparse, but also enforces the convolution between the filter and trained images to be sparse. The proposed model incorporates the strengths of both SC and MRFs, which enables it to represent general images with both generative and discriminative abilities. The resulting minimization is tackled by the combination of alternating optimization and Iterative Reweighted Least Square (IRLS). Experiments conducted on compressed sensing (CS) application show its great potentials both quantitatively and qualitatively.

Index Terms—Convolutional sparse coding, Analysis learning, MRFs, Deconvolutional network, CS recovery.

1. INTRODUCTION

One of the most successful approaches to solving inverse problems in image processing is to minimize a suitable energy functional whose minimizer provides a trade-off between a data consistency term and a prior distribution term. Early priori-derived algorithms include the total variation, wavelet sparsity or more general Markov random fields (MRFs) priors. The traditional predefined dictionaries may fail to sparsely represent the target images. To alleviate it, adaptive learning approaches (e.g., dictionary updating) have attracted remarkable attention over the years. Learning/data-driven approaches have become an increasingly important tool in machine learning and computer vision. Image features can be learned and subsequently used for various image processing applications [1, 2, 3]. Broadly speaking, there are two classes of learning strategies to model the image manifolds: synthesis learning models and analysis learning models.

Historically, sparse coding based representations refer to the so-called sparse synthesis model. It suggests that every image patch is composed of a linear combination of only a few atoms from dictionary [1, 2, 4, 5, 6]. The synthesis model, including global and specialized dictionary learning algorithms, has shown its promise in a series of applications [1, 2, 3, 4]. Nevertheless, the conventional patch-based sparse representation has a fundamental disadvantage that the important spatial structures of the image of interest may be lost due to the subdivision into patches that are independent of each other. To make up the deficiencies, Zeiler *et al.* [7] proposed a deconvolutional network, whose main tool is the convolutional implementation of sparse coding (CSC):

$$\min_{d,z} \frac{\lambda_1}{2}\left\| x-\sum_{k=1}^{K} d_k * z_k \right\|_2^2 + \sum_{k=1}^{K} |z_k|_1 \qquad (1)$$

where the first and the second terms represent the reconstruction error and the l_1-norm penalty respectively; λ_1 is a regularization parameter that controls the relative weight of the sparsity term; * denotes the 2D discrete convolution operator; and filters are restricted to have unit energy to avoid trivial solutions. In the convolutional decomposition procedure, the decomposition does not need to divide the entire image into overlapped patches, and can naturally utilize the consistency prior. Recently, CSC has been applied in many reconstruction tasks and feature-based methods [8, 9, 10, 11, 12, 13, 14, 15].

Another point of view to consider sparse representations is the analysis learning model, also called co-sparse analysis [16]. It usually assumes the image is locally highly incoherent, providing nearly-zero responses, with most of the elements of the filter set. The most classical and popular of this kind is the total variation (TV), whose atoms are the discretized first-order differential operators. Generally, the objective of a co-sparse analysis learning model is to pursue a linear operator A, such that the resulting coefficient vector Ax_l is expected to be sparse. The analysis model is far less investigated compared to the well-known synthesis model, but it has been gaining more and more attention in recent years [17, 18, 19, 20, 21, 22]. In [22], Chen *et al.* revealed that the analysis operator learning problem, which focuses on the modeling of small image patches, is tantamount to learning filters in the Product-of-Experts (PoE) model [23], ignoring the coherence between over-lapping patches, and thus misses global support during image reconstruction. Chen *et al.* showed that the global image-based analysis model is equivalent to Fields-of-Experts (FoE) model [24, 25], whose potential function is:

$$p(x,\Theta)=\frac{1}{Z(\Theta)}\prod_{k=1}^{K}\phi(d_k^T * x;\alpha_k) \qquad (2)$$

where $Z(\Theta)$ is the normalizing function; $\phi(x;\alpha)=(1+x^2/2)^{-\alpha}$ is an expert function; α_k is the parameter of the filter d_k. By means of the convolutional operator, the FoE model can be seen

This work was supported in part by the National NSFC under 61362001, 6166103, the international postdoctoral exchange fellowship program.

as an extension of the co-sparse analysis model from a patch-based formulation to an image-based formulation.

The exploitation of learning synthesis and analysis operators at the same framework has received some attention of researchers. In ref. [26], Zhu *et al.* proposed to classify the natural image patches into two types of atomic subspaces: explicit manifolds of low dimensions for structural primitives and implicit manifolds of high dimensions for stochastic textures. Then they showed that SC model and MRF model are good strategies for pursuing the explicit manifolds and implicit manifolds, respectively. One is a generative and expansion way and the other is a discriminative and reducing entropy way. Unfortunately, although there are many researches that yields impressive results by developing synthesis or analysis operators separately, little research has been conducted to answer the natural but fundamental question: does there exist a basis/filter set that can integrate the strengths of synthesis and analysis operators, and be able to represent general images with both generative and discriminative abilities? Since in real-world applications, many larger image patches contain both geometric structures and stochastic textures, it is definitely important to find such an operators set that can describe the general case. In this paper, we treat the synthesis and analysis operator learning constraints (associated with CSC and FoE priors, respectively) as peers and incorporate them into a unified variational framework. To the best of our knowledge, this is the first time that the synthesis and analysis convolutional filters are incorporated to a learning procedure and constrained to be the identical one, subsequently enabling the learned filter to have both feature representation and high-frequency properties.

2. PROPOSED ALGORITHM

2.1. Proposed SADN Model

The synthesis model CSC in Eq. (1) only requires the coefficient to be sparse. This strategy may not be accurate enough for faithfully representing the original image. Therefore, it is substantially essential to absorb some other complementary learning strategies such that the learned filters/atoms have better representation and discriminative ability. Thanks to the flexibility of the convolutional filter and motivated by the filter-based FoE model [25], we additionally enforce the convolution of the filter and trained images to be sparse. Consolidating the strengths of CSC and FoE models, the synthesis and analysis convolutional sparse coding is formulated as follows:

$$\min_{d,z} \frac{\lambda_1}{2}\left\| x-\sum_{k=1}^{K_1} d_k * z_k \right\|_2^2 + \sum_{k=1}^{K_1}\left|z_k\right|_1 + \frac{\lambda_2}{2}\sum_{k=1}^{K_1}\left|d_k^T * x\right|_1 \tag{3}$$

where λ_1 and λ_2 are two penalty parameters that balance the weights of synthesis priori in the first two terms and analysis sparsity priori in the last term, respectively. It can be observed that the formulation Eq. (3) merges the spirits in Eq. (1) and Eq. (2) in a unified fashion.

As stated in the deconvolutional network (DN), the single-layer model (3) can be stacked to form a hierarchy by treating the feature maps $z_{k,l}^i$ of layer l as input for layer $l+1$. Concretely, for all the ith test images, $i=1,2,\cdots,I$, the objective function C_l of synthesis and analysis deconvolutional network (SADN) at layer l is as follows:

$$C_l(d,z)=\frac{\lambda_1}{2}\sum_{i=1}^{I}\sum_{c=1}^{K_{l-1}}\left\|\sum_{k=1}^{K_l} g_{k,c}^l\left(z_{k,l}^i * d_{k,c}^l\right)-z_{c,l-1}^i\right\|_2^2+\sum_{i=1}^{I}\sum_{k=1}^{K_l}\left|z_{k,l}^i\right|_p + \frac{\lambda_2}{2}\sum_{i=1}^{I}\sum_{c=1}^{K_{l-1}}\sum_{k=1}^{K_l}\left|g_{k,c}^l\left(d_{k,c}^l\right)^T * z_{c,l-1}^i\right|_q \tag{4}$$

where $z_{c,l-1}^i$ and $z_{k,l}^i$ are the input and output feature maps. $g_{k,c}^l$ is the connectivity sign with 0 or 1 value. Particularly, $g_{k,c}^1 \equiv 1$ in the formulation (3). Intuitively, the inferring of the filters and feature maps is done in a layer-wise manner starting from the first layer where the input feature map $z_{c,0}^i = x_c^i$.

One demonstration of SADN filters learned from the Fruit dataset [7] is shown in Fig. 1. As can be observed in Fig. 1(a), the learned filters by FoE present some special structures. We can find high-frequency filters as well as high-order derivative filters along different directions. These filters make the analysis prior based model to be able to capture the structures in natural images that cannot be captured by using only the first derivatives as in the total variation based methods. On the other hand, In Fig. 1(b), the convolutional filter bank by CSC contains diverse features with a large range of orientations. Finally, by means of incorporating the FoE-like penalty to the CSC formulation, our results in Fig. 1(c) contain containing fewer dataset-specific features and meanwhile more high-frequency filters along different directions. In fact, SADN filters are not only be required to non-locally represent the data images, but also enforced by the FoE-like penalty to its local convolution with the images to be sparse. Therefore, the learned filter exhibits representation beyond edge primitives, and may contain second or higher order derivative-like features, such as edge intersections, parallelism and symmetry cues. Many high-frequency filters as well as derivative filters along different directions can be observed in the SADN filter, but are absent in the CSC filter.

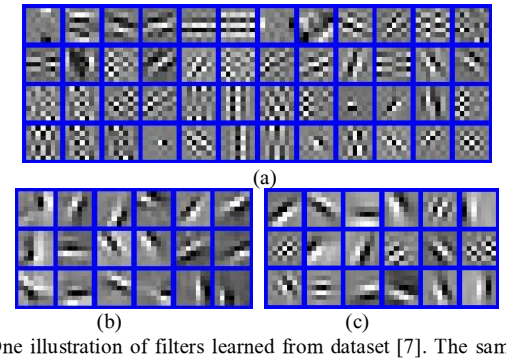


Fig. 1. One illustration of filters learned from dataset [7]. The sample filters learned with: (a) MRF filter [23], (b) DN filter, and (c) SADN filter.

Generally, jointly optimizing over the coefficient matrix and the dictionary/filter in the dictionary learning/filter learning is a non-convex problem, the alternating optimization procedure is usually used to approximate the global solution [1, 3, 4, 6, 27, 28]. Basically, these algorithms consist of alternatively learning the sparse approximation coefficient matrix when the dictionary/filter is considered fixed and then updating the dictionary with a fixed coefficient matrix. In this work, we also proceed to solve the kernel filter and coefficient in the SADN model via the alternative iterative scheme.

After removing the objective function that is irrelative to the variable d, the minimization in Eq. (4) with respect to d via the Iterative Reweighted Least Square (IRLS) is as follows:

$$\frac{\lambda_1}{2}\sum_{i=1}^{I}\sum_{c=1}^{K_{l-1}}\left\|\sum_{k=1}^{K_l} g_{k,c}^{l}(z_{k,l}^{i} * d_{k,c}^{l}) - z_{c,l-1}^{i}\right\|_2^2 + \frac{\lambda_2}{2}\sum_{i=1}^{I}\sum_{c=1}^{K_{l-1}}\sum_{k=1}^{K_l}\left|W_{i,c,k} g_{k,c}^{l}(d_{k,c}^{l})^T * z_{c,l-1}^{i}\right|_2^2 \tag{5}$$

where $W_{i,c,k} = 1/[g_{k,c}^{l}(d_{k,c}^{l})^T * z_{c,l-1}^{i}]$ denotes the weight obtained from the previous iteration. The minimization of Eq. (5) can be tackled by conjugate gradient (CG) solver.

Similarly, the partial objective function in Eq. (4) with regard to variable $z_{k,l}^{i}$ is written by the following formulation:

$$\frac{\lambda_1}{2}\sum_{i=1}^{I}\sum_{c=1}^{K_{l-1}}\left\|\sum_{k=1}^{K_l} g_{k,c}^{l}(z_{k,l}^{i} * d_{k,c}^{l}) - z_{c,l-1}^{i}\right\|_2^2 + \sum_{i=1}^{I}\sum_{k=1}^{K_l}\left|z_{k,l}^{i}\right|_p \tag{6}$$

Same as in DN, we still use the half-quadratic splitting technique to solve it as follow:

$$\frac{\lambda_1}{2}\sum_{i=1}^{I}\sum_{c=1}^{K_{l-1}}\left\|\sum_{k=1}^{K_l} g_{k,c}^{l}(z_{k,l}^{i} * d_{k,c}^{l}) - z_{c,l-1}^{i}\right\|_2^2 + \frac{\beta}{2}\sum_{i=1}^{I}\sum_{c=1}^{K_l}\left\|z_{k,l}^{i} - y_{k,l}^{i}\right\|_2^2 + \sum_{i=1}^{I}\sum_{k=1}^{K_l}\left|y_{k,l}^{i}\right|_p \tag{7}$$

where β is the continuation parameter. The introducing of auxiliary variables $y_{k,l}^{i}$ decouples the difficulty in the least square and the non-convex quasi-norm term. Minimizing Eq. (7) for a fixed β is performed by alternating between two steps: solve $y_{k,l}^{i}$ given values of $z_{k,l}^{i}$ and vice-versa.

2.2. CS Recovery

In the scenario of CS recovery, the observed data is modeled as $f = Mx + n$, where n is the Gaussian noise. Specifically, M denotes the incomplete Fourier encoding matrix in MRI reconstruction. Dictionary learning for medical imaging has received a lot of attention. The early learning strategy was typically done using reference images [29, 30]. For instance, Chen *et al.* [30] learned a patch-based K-SVD dictionary from fully sampled reference images, and employed the norm for sparsity in the reconstruction. The results showed improved performance with the learned dictionary over CSMRI with wavelets as sparsifying transform. However, all these improvements are limited, indicating that a dictionary learned from a reference image would not be able to effectively sparsify new features in the current scan/measurement. In [31, 3], the authors proposed to use adaptive sparsifying transforms instead of fixed ones learned from dataset. The idea is to simultaneously learn the sparsifying transform and reconstruct the image. The experimental results showed significant improvements over previous algorithms.

After the SADN filter kernels are learned from the reference dataset, we need to solve the recovery image. For multi-layer SADN model, the updating procedure is more complicated. For simplification, currently we only consider the one-layer model for CS recovery, denoted as SADN-1. i.e. $K_0 = 1$, $g_{k,1}^{1} \equiv 1$, then the recovery of x from f can be achieved via solving:

$$\min_{x,z} v/2\left\|f - Mx\right\|_2^2 + \frac{\lambda_1}{2}\left\|x - \sum_{k=1}^{K_1} d_k * z_k\right\|_2^2 + \sum_{k=1}^{K_1}\left|z_k\right|_p + \frac{\lambda_2}{2}\sum_{k=1}^{K_1}\left|d_k^T * x\right|_q \tag{8}$$

where the weight v is set as $v = \lambda/\sigma$, λ is a positive constant. Eq. (8) can be addressed by the alternative iterative scheme:

$$\begin{aligned} x^i &= \arg\min_x v/2\left\|f - Mx\right\|_2^2 + \frac{\lambda_1}{2}\left\|x - \sum_{k=1}^{K_1} d_k * z_k\right\|_2^2 + \frac{\lambda_2}{2}\sum_{k=1}^{K_1}\left|d_k^T * x\right|_q \\ z^i &= \arg\min_z \frac{\lambda_1}{2}\left\|x - \sum_{k=1}^{K_1} d_k * z_k\right\|_2^2 + \sum_{k=1}^{K_1}\left|z_k\right|_p \end{aligned} \tag{9}$$

For the x-subproblem, currently we simply solve it via the IRLS technique. It is worth noting that our scheme could be improved by using the Half-quadratic minimization with continuation like in [32].

3. EXPERIMENTAL RESULTS

In this section, we first illustrate some filter properties of the SADN, then its CS recovery performance is evaluated. The peak signal-to-noise ratio (PSNR) and high-frequency error norm (HFEN) are used for quantitative comparison.

The trend of visual effect of the learned filters with respect to parameter λ_2/λ_1 is shown in Fig. 2. As can be expected, with a larger λ_2/λ_1, more penalty will be concentrated on the analysis term, thus the generated filter appears closer to the analysis penalty. In order to balance the feature representation generative and discriminative ability, we empirically set $\lambda_2/\lambda_1 = 3\times10^{-3}$. To investigate the influence of the training dataset, we retrain our model by using (i) the Fruit dataset as training dataset1 and (ii) the City dataset as training dataset2 in ref. [7] for comparison. All the images in dataset1 are fruits such as bananas and apples. Instead, the images from dataset2 contains face, leaves, hat, etc. Two observations are found from Fig. 3. First, SADN filters are more diverse, verifying that the additional analysis operator regularization is beneficial to recover images with irregular patterns. Second, comparing the results of training dataset1 with those of dataset2, it can be found that the filters obtained by DN are more distinct than that by SADN. That is, the dependence of SADN on training dataset is less than that of DN. Therefore, the SADN is more robust to training datasets.

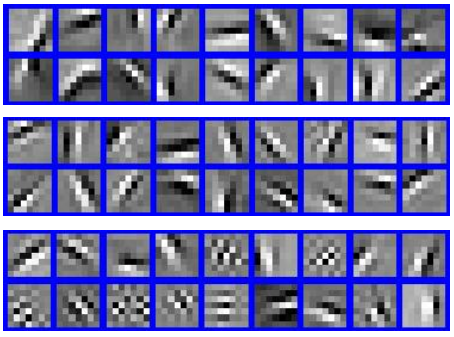


Fig. 2. Filter kernels learned under different values of parameter λ_2/λ_1.

In CS-recovery experiment, the present method is compared with the three patch-based sparse representation approaches: DLMRI [31], Grad-DL [33], and NLR-CS [34]. Both DLMRI and Grad-DL method utilize the popular K-SVD dictionary learning to represent the underlying data. DLMRI is devoted to represent the image while Grad-DL is committed to employ the dictionary learning in the sparse image gradients domain. NLR-CS utilizes the non-convex logarithm function to

approximate the rank penalty of grouped patches, pursuing a structured sparse representation.

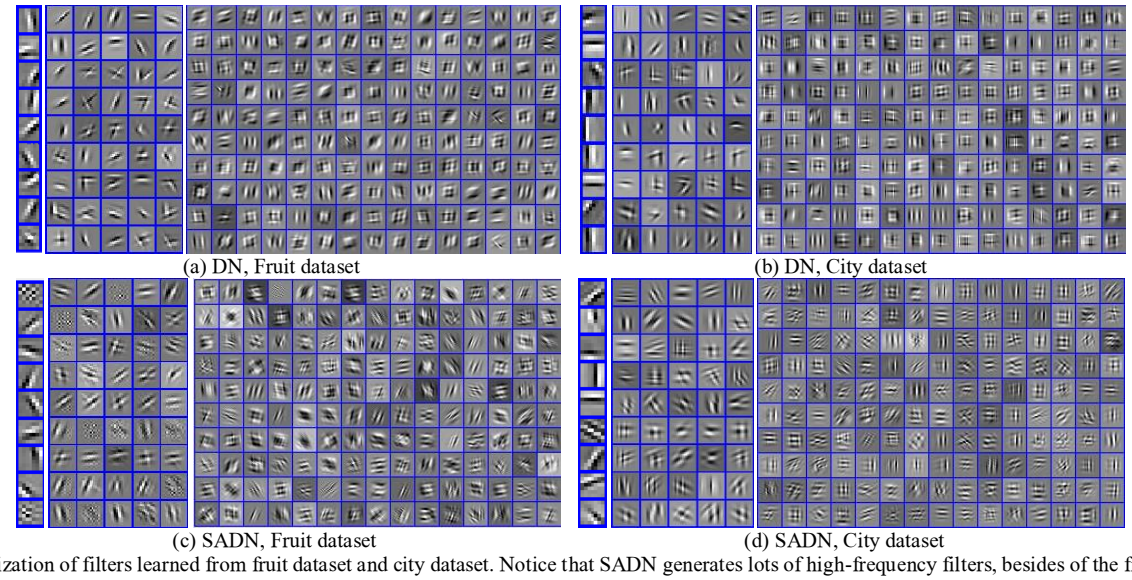
Table 1 shows the PSNR and HFEN comparison results under 2D random and pseudo radial under-sampling schemes. It can be observed that NLR-CS achieves higher PSNR values than the former two dictionary learning methods and also better than SADN-1 sometimes. However, the HFEN values gained by NLR-CS are worse than those of SADN-1 in the 2D random pattern case, implying that NLR-CS produces over-smooth results. Moreover, it is worth noting that NLR-CS employs sophisticated techniques such as good initialization estimation and warmstart for better approximating the non-convex problem. Some visual recovery results are shown in Fig. 4.

4. CONCLUSIONS

In this paper, a synthesis and analysis simultaneously constrained convolutional sparse coding model has been presented. In the model, taking advantage of the flexibility of convolution operator, the relation between the analysis dictionary and synthesis operator was built. The joint generative-discriminative filter learning is implemented by alternating optimization and IRLS solver. The effectiveness of the generative-discriminative criteria is demonstrated in case of CS recovery. This work sheds a new insight to how synthesis and analysis convolutional filters work from the variational and manifold perspectives.

Table 1. Recovery PSNR/HFEN values on three images [35] (T2axialbrain, KneeMRISag, COW0001) at undersampling rates of 63%, 73% and 80%.

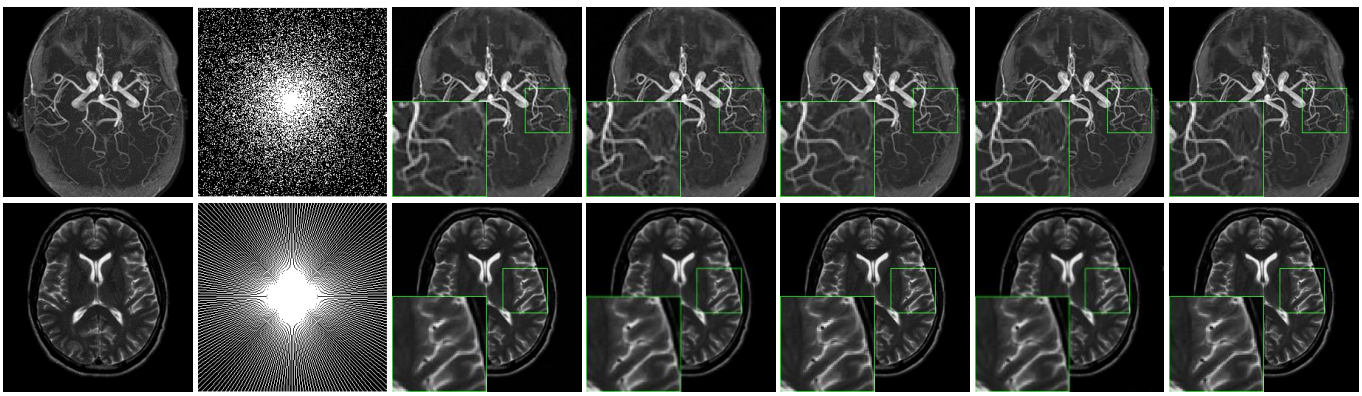
	DLMRI	Grad-DL	NLR-CS	DN-1, $p=0.8$	SADN-1, $p=0.8$
2D Random	38.52/0.327	43.77/0.188	50.44/0.120	51.20/0.043	**51.47/0.041**
	36.77/0.434	40.84/0.280	47.09/0.233	46.93/0.083	**47.22/0.078**
	35.22/0.575	38.42/0.407	44.32/0.682	43.39/0.146	43.94/**0.145**
	34.87/0.534	36.60/0.484	39.18/0.567	36.68/0.488	37.69/**0.423**
	32.62/0.781	33.92/0.721	36.34/0.849	33.99/0.773	34.89/**0.641**
	31.22/0.920	32.07/0.914	34.78/1.027	32.40/0.950	33.07/**0.807**
	34.88/0.518	37.92/0.372	42.98/0.404	42.46/0.164	**42.98/0.163**
	32.89/0.727	35.11/0.581	39.05/0.675	38.70/0.314	**39.34/0.302**
	31.56/0.906	33.28/0.754	37.15/0.828	36.25/0.473	36.82/**0.445**
Pseudo Radial	36.68/0.466	39.54/0.359	44.57/0.150	44.37/0.133	**44.80/0.127**
	34.87/0.695	36.65/0.605	41.21/0.350	39.83/0.389	40.46/0.361
	32.75/1.059	33.88/1.015	38.41/0.553	36.16/0.770	36.79/0.707
	32.88/0.708	33.02/0.763	34.57/0.589	32.97/0.812	33.48/0.732
	30.52/1.220	30.44/1.322	32.12/1.030	30.31/1.465	30.69/1.282
	28.41/1.806	28.09/1.973	29.78/1.560	27.91/2.216	28.34/1.945
	32.21/0.884	32.89/0.874	37.07/0.511	36.31/0.516	36.78/**0.491**
	30.39/1.297	30.98/1.270	33.67/0.907	32.82/1.031	33.21/0.982
	28.47/1.843	28.63/1.889	31.35/1.366	30.41/1.580	30.63/1.511



(a) DN, Fruit dataset (b) DN, City dataset

(c) SADN, Fruit dataset (d) SADN, City dataset

Fig. 3. Visualization of filters learned from fruit dataset and city dataset. Notice that SADN generates lots of high-frequency filters, besides of the first derivatives.



(a) (b) (c) (d) (e) (f) (g)

Fig.4. CS recovered MR images with 73% 2D random (Top line) and 63% pseudo radial (Bottom line) under-sampling scheme. (a) Reference image, (b) sampling pattern, (c) DLMRI, (d) Grad-DL, (e) NLR-CS, (f) DN-1 and (g) SADN-1.

5. REFERENCES

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

[2] J. Mairal, M. Elad, and G. Sapiro, "Sparse representation for color image restoration," *IEEE Trans. Image Process.*, vol. 17, no. 1, pp. 53-69, 2008.

[3] Q. Liu, S. Wang, K. Yang, J. Luo, Y. Zhu, and D. Liang, "Highly undersampled magnetic resonance image reconstruction using two-level Bregman method with dictionary updating," *IEEE Trans. Med. Imaging*, vol. 32, no. 7, pp. 1290-1301, 2013.

[4] Q. Liu, D. Liang, Y. Song, J. Luo, Y. Zhu and W. Li, "Augmented Lagrangian-based sparse representation method with dictionary updating for image deblurring," *SIAM J. Imag. Sci.*, vol. 6, no. 3, pp. 1689-1718, 2013.

[5] H. Lee, A. Battle, R. Raina, and A.Y. Ng, "Efficient sparse coding algorithms," in *NIPS*, pp. 801-808, 2006.

[6] Q. Liu, J. Luo, S. Wang, M. Xiao and M. Ye, "An augmented Lagrangian multi-scale dictionary learning algorithm," *EURASIP J. Adv. Signal Process.*, vol. 58, no. 1, pp. 1-16, 2011.

[7] M.D. Zeiler, D. Krishnan, G.W. Taylor, and R. Fergus, "Deconvolutional networks," in *Proc. IEEE Conf. CVPR*, pp. 2528-2535, 2010.

[8] S. Gu, W. Zuo, Q. Xie, et al, "Convolutional sparse coding for image super-resolution," in *Proc. IEEE Conf. ICCV*, pp. 1823-1831, 2015.

[9] D. Carrera, G. Boracchi, A. Foi, B. Wohlberg, "Detecting anomalous structures by convolutional sparse models," *International Joint Conference on Neural Networks (IJCNN)*.

[10] Y. Zhu and S. Lucey, "Convolutional sparse coding for trajectory reconstruction", *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 37, no. 3, pp. 529-540, 2015.

[11] A. Serrano, F. Heide, D. Gutierrez, "Convolutional sparse coding for high dynamic range imaging," *Computer Graphics*, vol. 35, no. 2, 2016.

[12] F. Heide, L. Xiao, A. Kolb, M. Hullin, W. Heidrich, "Imaging in scattering media using correlation image sensors and sparse convolutional coding," *Optics Express*, 2014.

[13] M.D. Zeiler, G.W. Taylor, and R. Fergus, "Adaptive deconvolutional networks for mid and high level feature learning," in *Proc. IEEE Conf. ICCV*, pp. 2018-2025, 2011.

[14] K. Kavukcuoglu, P. Sermanet, Y.L. Boureau, K. Gregor, M. Mathieu and Y. LeCun, "Learning convolutional feature hierarchies for visual recognition," in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 23, pp. 1090-1098, 2010.

[15] B. Chen, G. Polatkan, G. Sapiro, D. Blei, D. Dunson, and L. Carin, "Deep learning with hierarchical convolutional factor analysis," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol.35, no.8, pp.1887-1901, 2013.

[16] M. Elad, P. Milanfar, and R. Rubinstein, "Analysis versus synthesis in signal priors," *Inverse Problems*, 23(3):947-968, 2007.

[17] S. Hawe, M. Kleinsteuber, and K. Diepold, "Analysis operator learning and its application to image reconstruction," *IEEE Trans. Image Process.*, vol. 22, no. 6, pp. 2138-2150, 2013.

[18] G. Peyr´e and J. Fadili, "Learning analysis sparsity priors," In *Proc. of Sampta'11*, 2011.

[19] R. Rubinstein, T. Peleg, and M. Elad, "Analysis K-SVD: A dictionary learning algorithm for the analysis sparse model," *IEEE Trans. Signal Process.*, vol. 61, no. 3, pp. 661-677, 2013.

[20] M. Yaghoobi, S. Nam, R. Gribonval, and M.E. Davies, "Analysis operator learning for over-complete co-sparse representations," In *European Signal Processing Conference*, 2011.

[21] M. Yaghoobi, S. Nam, R. Gribonval, and M.E. Davies, "Noise aware analysis operator learning for approximately cosparse signals," In *ICASSP*, 2012.

[22] Y. Chen, R. Ranftl, and T. Pock, "Insights into analysis operator learning: From patch-based sparse models to higher order MRFs," *IEEE Trans. Image Process.*, vol. 23, no. 3, pp. 1060-1072, 2014.

[23] G. E. Hinton, "Products of experts," In *ICANN*, pp. 1-6, 1999.

[24] S. Roth and M. J. Black, "Fields of experts: A framework for learning image priors," in *Proc. IEEE Conf. CVPR*, vol. 2, San Diego, California, Jun. 2005, pp. 860–867.

[25] S. Roth and M.J. Black, "Fields of experts," *Int J. Comput. Vis.*, vol. 82, no. 2, pp. 205-229, 2009.

[26] S. Zhu, K. Shi and Z. Si, "Learning explicit and implicit visual manifolds by information projection," *Pattern Recognit. Lett.*, vol. 31, no. 8, pp. 667-685, 2010.

[27] F. Heide, W. Heidrich, G. Wetzstein, "Fast and flexible convolutional sparse coding," in *Proc. IEEE Conf. CVPR*, 2015, pp. 5135-5143.

[28] B. Wohlberg, "Efficient algorithms for convolutional sparse representations," *IEEE Trans. Image Process.*, vol. 25, no. 1, pp. 301-315, 2016.

[29] I. Tosic, I. Jovanovic, P. Frossard, M. Vetterli, and N. Duric, "Ultrasound tomography with learned dictionaries," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process.*, 2010, pp. 5502–5505.

[30] Y. Chen, X. Ye, and F. Huang, "A novel method and fast algorithm for MR image reconstruction with significantly under-sampled data," *Inverse Probl. Imag.*, vol. 4, no. 2, pp. 223–240, 2010.

[31] 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.

[32] J. Yang, Y. Zhang and W. Yin, "A fast alternating direction method for TVL1-L2 signal reconstruction from partial Fourier data," *IEEE J. Sel. Topics Signal Process*, vol. 4, no. 2, pp. 288–297, 2010.

[33] Q. Liu, S. Wang, L. Ying, X. Peng, Y. Zhu, and D. Liang, "Adaptive dictionary learning in sparse gradient domain for image recovery," *IEEE Trans. Image Process,* vol. 22, no. 12, pp. 4652-4663, 2013.

[34] W. Dong, G. Shi, X. Li, Y. Ma, and F. Huang, "Compressive sensing via nonlocal low-rank regularization," *IEEE Trans. Image Process,* vol. 23, no. 8, pp. 3618-3632, 2014.

[35] Am. Radiol. Services [Online]. Available: http://www3.americanradiology. com/pls/web1/wwmain.home