Multi-layer Residual Sparsifying Transform (MARS) Model for Low-dose CT Image Reconstruction

Xikai Yang[1], Yong Long[1], Saiprasad Ravishankar[2]
[1]University of Michigan - Shanghai Jiao Tong University Joint Institute,
Shanghai Jiao Tong University, Shanghai 200240, China
[2]Department of Computational Mathematics, Science and Engineering
and Department of Biomedical Engineering,
Michigan State University, East Lansing, MI 48824, USA
Version typeset May 28, 2021

Author to whom correspondence should be addressed. Mailing address: Room 402, Longbin Building, Shanghai Jiaotong University, Shanghai, China. Electronic mail: yong.long@sjtu.edu.cn

Abstract

Purpose: Signal models based on sparse representations have received considerable attention in recent years. On the other hand, deep models consisting of a cascade of functional layers, commonly known as deep neural networks, have been highly successful for the task of object classification and have been recently introduced to image reconstruction. In this work, we develop a new image reconstruction approach based on a novel multi-layer model learned in an unsupervised manner by combining both sparse representations and deep models. The proposed framework extends the classical sparsifying transform model for images to a Multi-lAyer Residual Sparsifying transform (MARS) model, wherein the transform domain data are jointly sparsified over layers. We investigate the application of MARS models learned from limited regular-dose images for low-dose CT reconstruction using Penalized Weighted Least Squares (PWLS) optimization.
Methods: We propose new formulations for multi-layer transform learning and image reconstruction. We derive an efficient block coordinate descent algorithm to learn the transforms across layers, in an unsupervised manner from limited regular-dose images. The learned model is then incorporated into the low-dose image reconstruction phase.
Results: Low-dose CT experimental results with both the XCAT phantom and Mayo Clinic data show that the MARS model outperforms conventional methods such as FBP and PWLS methods based on the edge-preserving (EP) regularizer in terms of two numerical metrics (RMSE and SSIM) and noise suppression. Compared with the single-layer learned transform (ST) model, the MARS model performs better in maintaining some subtle details.
Conclusions: This work presents a novel data-driven regularization framework for CT image reconstruction that exploits learned multi-layer or cascaded residual sparsifying transforms. The image model is learned in an unsupervised manner from limited images. Our experimental results demonstrate the promising performance of the proposed multi-layer scheme over single-layer learned sparsifying transforms. Learned MARS models also offer better image quality than typical nonadaptive PWLS methods.

I. Introduction

Signal models exploiting sparsity have been shown to be useful in a variety of of imaging and image processing applications such as compression, restoration, denoising, reconstruction, etc.[1,2,3,4] Natural signals can be modeled as sparse in a synthesis dictionary (i.e., represented as a linear combinations of a few dictionary atoms or columns) or in a sparsifying transform domain. Transforms such as wavelets[5] and the discrete cosine transform (DCT) are well-known to sparsify images. Synthesis dictionary learning[6] and analysis dictionary learning[7] methods adapt such models to data and involve algorithms such as K-SVD[7], the Chasing Butterflies approach[8], and some others. The underlying dictionary learning problems are typically NP-hard and the corresponding algorithms often involve computationally expensive updates that limit their applicability to large-scale data. In contrast, the recently proposed sparsifying transform learning approaches[9] involve exact and highly efficient updates in the algorithms. In particular, the transform model suggests that the signal is approximately sparse in a transformed domain. Furthermore, Ravishankar *et al*[10,11,12] demonstrated the applicability of adaptive sparsifying transforms for several applications such as image denoising and medical image reconstruction.

On the other hand, deep models with nested network structure popularly known as deep neural networks provide remarkable results for classification and regression across various fields[13]. Given a task-based loss function for network parameter estimation, algorithms based on gradient back-propagation sequentially reduce the error between a known target (ground truth) and the network prediction. Another approach from a few research groups combines deep network architectures with probabilistic models during learning, and this generative Bayesian model[14] attains a superior performance during the inference process. Morever, the connections between sparse modeling and deep neural networks has also been exploited. For example, the multi-layer convolutional (synthesis) sparse coding model[15,16] provides a new interpretation of convolutional neural networks (CNNs), where the pursuit of sparse representation from a given input signal complies with the forward pass in a CNN. In the meantime, multi-layer sparsifying transforms make the most direct connection with CNNs in the model and enable sparsifying an input image successively over layers[17], creating a rich and more complete sparsity model, whose learning in an unsupervised manner and from limited data also forms the core of this work.

One of the most important applications of such image models is for medical image reconstruction. In particular, an important problem in X-ray computed tomography (CT) is reducing the X-ray exposure to patients while maintaining good image reconstruction quality. A conventional method for CT reconstruction is the analytical filtered back-projection (FBP)[18]. However, image quality degrades severely for FBP when the radiation dose is reduced. In contrast, model-based image reconstruction (MBIR) exploits CT forward models and statistical models together with image priors to achieve often better image quality[19].

A typical MBIR method for low-dose CT (LDCT) is the penalized weighted least squares (PWLS) approach. The cost function for PWLS includes a weighted quadratic data-fidelity term and a penalty term or regularizer capturing prior information or model of the object[20,21,22]. Recent works have shown promising LDCT reconstruction quality by incorporating data-driven models into the regularizer, where the models are learned from datasets of images or image patches. In particular, PWLS reconstruction with adaptive sparsifying transform-based regularization has shown promise for tomographic reconstruction[23,24,25,26,27]. Recent work has also shown that they may generalize better to unseen new data than supervised deep learning schemes[28]. The adaptive transform-based image reconstruction algorithms can exploit a variety of image models[23,26,29] learned in an unsupervised manner from limited training images, and involve efficient closed-form solutions for sparse coding.

In this work, we propose a new formulation and algorithm for learning a multi-layer transform model[17], where the transform domain residuals (the difference between transformed data and their sparse approximations) are successively sparsified over several layers. We refer to the model as the Multi-lAyer Residual Sparsifying transform (MARS) model. The transforms are learned over several layers from images to jointly minimize the transform domain residuals across layers, while enforcing sparsity conditions in each layer. Importantly, the filters beyond the first layer can help better exploit finer features (e.g., edges and correlations) in the residual maps. We investigate the performance of unsupervised learning of MARS models from limited data for LDCT reconstruction using PWLS. We propose efficient block coordinate descent algorithms for both learning and reconstruction. Experimental results with the XCAT phantom and Mayo Clinic data illustrate that the learned MARS model outperforms conventional methods such as FBP and PWLS methods based on the non-adaptive edge-preserving (EP) regularizer in terms of two numerical metrics (RMSE and SSIM) and noise suppression. Compared with the recent learned single-layer transform model, the MARS model performs better in maintaining some subtle details.

In the following sections, we will first study how to train our proposed model in detail in Section II, where we will discuss the corresponding problem formulations in Section II-A, followed by our algorithms in Section II-B. The experimental results with both the XCAT phantom and Mayo Clinic data are presented in Section III. Section IV presents a discussion of the proposed methods and results and concludes.

II. Methods

II.A. Formulations for MARS Training and LDCT reconstruction

Here, we introduce the proposed general multi-layer transform learning framework and the formulation for LDCT image reconstruction. Fig. 1 illustrates the structure of our multi-layer residual sparsifying transform model, where $\mathbf{\Omega}_l$ denotes the transform in the lth layer. These transforms capture higher order image information by sparsifying the transform domain residual maps layer by layer. The MARS learning cost and constraints are shown in Problem (P0), which is an extension of simple single-layer transform learning[9,17].

$$\begin{aligned} &\min_{\{\mathbf{\Omega}_l, \mathbf{Z}_l\}} \sum_{l=1}^{L} \left\{ \|\mathbf{\Omega}_l \mathbf{R}_l - \mathbf{Z}_l\|_F^2 + \eta_l^2 \|\mathbf{Z}_l\|_0 \right\}, \\ &\text{s.t.} \quad \mathbf{R}_l = \mathbf{\Omega}_{l-1}\mathbf{R}_{l-1} - \mathbf{Z}_{l-1}, 2 \le l \le L, \ \ \mathbf{\Omega}_l^T \mathbf{\Omega}_l = \mathbf{I}, \forall l. \end{aligned} \tag{P0}$$

Here, $\{\mathbf{\Omega}_l \in \mathbb{R}^{p\times p}\}$ and $\{\mathbf{Z}_l \in \mathbb{R}^{p\times N}\}$ denote the sets of learned transforms and sparse coefficient maps, respectively, for the $1 \le l \le L$ layers and "F" denotes the Frobenius norm. The total number of training patches is denoted by N. Parameter η_l controls the maximum allowed sparsity level (computed using the ℓ_0 "norm" penalty) for $\mathbf{Z}_l$. The residual maps $\{\mathbf{R}_l \in \mathbb{R}^{p\times N}\}$ are defined in recursive form over layers, with $\mathbf{R}_1$ denoting the input training data. We assume $\mathbf{R}_1$ to be a matrix, whose columns are (vectorized) patches drawn from image data sets. The unitary constraint for each $\mathbf{\Omega}_l$ enables closed-form solutions for the sparse coefficient and transform update steps in our algorithms. The MARS model learned via (P0) can then be used to construct a data-driven regularizer in PWLS as shown in Problem (P1).

$$\min_{\mathbf{x}\ge\mathbf{0}} \frac{1}{2}\|\mathbf{y} - \mathbf{A}\mathbf{x}\|_{\mathbf{W}}^2 + \beta \mathsf{S}(\mathbf{x}), \tag{P1}$$

$$\mathsf{S}(\mathbf{x}) \triangleq \min_{\{\mathbf{Z}_l\}} \sum_{l=1}^{L} \left\{ \|\boldsymbol{\Omega}_l \mathbf{R}_l - \mathbf{Z}_l\|_F^2 + \gamma_l^2 \|\mathbf{Z}_l\|_0 \right\},$$
$$\text{s.t. } \mathbf{R}_l = \boldsymbol{\Omega}_{l-1}\mathbf{R}_{l-1} - \mathbf{Z}_{l-1}, 2 \leq l \leq L\,, \mathbf{R}_1^j = \mathbf{P}^j \mathbf{x},\ \forall\, j.$$

In particular, we reconstruct the image $\mathbf{x} \in \mathbb{R}^{N_p}$ from noisy sinogram data $\mathbf{y} \in \mathbb{R}^{N_d}$ by solving (P1), where N_p denotes the number of pixels. $\mathbf{A} \in \mathbb{R}^{N_d \times N_p}$ is the system matrix of the CT scan and $\mathbf{W} = \mathsf{diag}\{w_i\} \in \mathbb{R}^{N_d \times N_d}$ is the diagonal weighting matrix with elements being the estimated inverse variance of y_i. Operator $\mathbf{P}^j \in \mathbb{R}^{p \times N_p}$ extracts and vectorizes the jth patch of $\mathbf{x}$ as $\mathbf{P}^j\mathbf{x}$. Overlapping image patches are extracted with appropriate patch stride (1 pixel stride in our experiments). The jth columns of $\mathbf{R}_l$ and $\mathbf{Z}_l$ are denoted $\mathbf{Z}_l^j$ and $\mathbf{R}_l^j$, respectively. The non-negative parameters $\{\gamma_l\}$ control the sparsity of the coefficient maps in different layers, and $\beta > 0$ captures the relative trade-off between the data-fidelity term and regularizer.

II.B. Algorithms for Learning and Reconstruction

Fig. 2 provides an overview of the proposed method. The whole algorithm is divided into two stages: training and reconstruction. In the training stage, we solve (P0) using a block coordinate descent (BCD) method to learn a multi-layer sparsifying transform model in an unsupervised manner from (unpaired) regular-dose images. For the reconstruction stage, the prior information incorporated into learned transform would be designed into regularizer term, and iterative algorithm accomplishes the reconstruction for the CT image as we will show in the later section.

II.B.1. MARS Learning Algorithm

We propose an exact block coordinate descent (BCD) algorithm for the nonconvex Problem (P0) that cycles over updating $\mathbf{Z}_l$ (*sparse coding step*) followed by updating the corresponding $\boldsymbol{\Omega}_l$ (*transform update step*) for $1 \leq l \leq L$. The algorithmic details are shown in **Algorithm 1**. In each step, the remainder of the variables (that are not optimized) are kept fixed. The BCD algorithm provides a very efficient way to minimize the cost function and is shown to empirically work well with appropriate initialization. Recent works involving transform learning[28,30] have shown that such efficient alternating minimization or BCD algorithms can provably converge to the critical points of the underlying problems. In particular, we show that under the unitarity condition on the transforms, every subproblem in the block coordinate descent minimization approach can be solved exactly. We initialize the algorithm with the 2D DCT for $\mathbf{\Omega}_1$ and the identity matrices for $\{\mathbf{\Omega}_l\}_{l=2}^{L}$, respectively. The initial $\{\mathbf{Z}_l\}$ are all-zero matrices.

Since the residuals are defined recursively in (P0), for the sake of simplicity of the algorithmic description, we first define matrices $\mathbf{B}_p^q (p < q)$, which can be regarded as back-propagation matrices from the qth to pth layers.

$$\begin{aligned}\mathbf{B}_p^q &= \mathbf{\Omega}_{p+1}^T \mathbf{Z}_{p+1} + \mathbf{\Omega}_{p+1}^T \mathbf{\Omega}_{p+2}^T \mathbf{Z}_{p+2} + ... + \mathbf{\Omega}_{p+1}^T \mathbf{\Omega}_{p+2}^T ... \mathbf{\Omega}_q^T \mathbf{Z}_q \\ &= \sum_{k=p+1}^{q} \Big(\prod_{s=p+1}^{k} \mathbf{\Omega}_s^T \Big) \mathbf{Z}_k.\end{aligned} \tag{1}$$

(a) Sparse Coding Step for $\mathbf{Z}_l$

Here, we solve (P0) for $\mathbf{Z}_l$ with all other variables fixed. The corresponding nonconvex subproblem is as follows:

$$\min_{\mathbf{Z}_l} \sum_{i=l}^{L} \Big\{ \|\mathbf{\Omega}_i \mathbf{R}_i - \mathbf{Z}_i\|_F^2 \Big\} + \eta_l^2 \|\mathbf{Z}_l\|_0. \tag{2}$$

Using the definitions of the residual matrices and the backpropagation matrices $\mathbf{B}_p^q\,(p < q)$ along with the unitary property of the transforms allows us to rewrite (2) as:

$$\min_{\mathbf{Z}_l} \|\mathbf{Z}_l - \mathbf{\Omega}_l \mathbf{R}_l\|_F^2 + \sum_{i=l+1}^{L} \|\mathbf{Z}_l + \mathbf{B}_l^i - \mathbf{\Omega}_l \mathbf{R}_l\|_F^2 + \eta_l^2 \|\mathbf{Z}_l\|_0. \tag{3}$$

We can now rewrite subproblem (3) as $\min_{\mathbf{Z}_l} (L-l+1) \times \|\mathbf{Z}_l - (\mathbf{\Omega}_l \mathbf{R}_l - \frac{1}{L-l+1} \sum_{i=l+1}^{L} \mathbf{B}_l^i)\|_F^2 + \eta_l^2 \|\mathbf{Z}_l\|_0$. This problem has a similar form as the single-transform sparse coding problem[9], and the optimal solution $\hat{\mathbf{Z}}_l$ is obtained as in (4), where $H_\eta(\cdot)$ denotes the *hard-thresholding* operator that sets elements with magnitude less than the threshold η to zero.

$$\hat{\mathbf{Z}}_l = \begin{cases} H_{\eta_l/\sqrt{L-l+1}}\Big(\mathbf{\Omega}_l \mathbf{R}_l - \frac{1}{L-l+1} \sum_{i=l+1}^{L} \mathbf{B}_l^i \Big), & 1 \le l \le L-1, \\ H_{\eta_L}(\mathbf{\Omega}_L \mathbf{R}_L), & l = L. \end{cases} \tag{4}$$

(b) Transform Update Step for $\mathbf{\Omega}_l$

Here, we fix $\{\mathbf{Z}_l\}$ and all $\mathbf{\Omega}_j$ (except the target $\mathbf{\Omega}_l$ in (P0)) and solve the following subproblem:

$$\min_{\mathbf{\Omega}_l} \sum_{i=l}^{L} \left\{ \|\mathbf{\Omega}_i \mathbf{R}_i - \mathbf{Z}_i\|_F^2 \right\} \quad \text{s.t.} \quad \mathbf{\Omega}_l^T \mathbf{\Omega}_l = \mathbf{I}. \tag{5}$$

Similar to (3), we rewrite (5) using the backpropagation matrices $\mathbf{B}_p^q \, (p < q)$ as follows:

$$\begin{aligned} &\min_{\mathbf{\Omega}_l : \mathbf{\Omega}_l^T \mathbf{\Omega}_l = \mathbf{I}} \|\mathbf{\Omega}_l \mathbf{R}_l - \mathbf{Z}_l\|_F^2 + \sum_{i=l+1}^{L} \|\mathbf{\Omega}_l \mathbf{R}_l - \mathbf{Z}_l - \mathbf{B}_l^i\|_F^2, \\ \sim &\min_{\mathbf{\Omega}_l : \mathbf{\Omega}_l^T \mathbf{\Omega}_l = \mathbf{I}} (L - l + 1) \times \left\| \mathbf{\Omega}_l \mathbf{R}_l - \mathbf{Z}_l - \frac{1}{L - l + 1} \sum_{i=l+1}^{L} \mathbf{B}_l^i \right\|_F^2, \end{aligned} \tag{6}$$

where the last relation (equality) holds up to an additive term that is independent of $\mathbf{\Omega}_l$. We can obtain a solution to (6) by exploiting the unitarity of $\mathbf{\Omega}_l$. First, denoting the full singular value decomposition (SVD) of the matrix $\mathbf{G}_l$ below by $\mathbf{U}_l \mathbf{\Sigma}_l \mathbf{V}_l^T$, the optimal solution to (6) is as (8).

$$\mathbf{G}_l = \begin{cases} \mathbf{R}_l \left(\mathbf{Z}_l + \frac{1}{L-l+1} \sum_{i=l+1}^{L} \mathbf{B}_l^i \right)^T, & 1 \le l \le L - 1, \\ \mathbf{R}_L \mathbf{Z}_L^T, & l = L. \end{cases} \tag{7}$$

$$\hat{\mathbf{\Omega}}_l = \mathbf{V}_l \mathbf{U}_l^T \tag{8}$$

Algorithm 1 MARS Learning Algorithm

Input: training data $\mathbf{R}_1$, all-zero initial $\{\tilde{\mathbf{Z}}_l^{(0)}\}$, initial $\tilde{\mathbf{\Omega}}_1^{(0)} = $ 2D DCT, identity matrices for initial $\{\tilde{\mathbf{\Omega}}_l^{(0)}\}_{l=2}^{L}$, thresholds $\{\eta_l\}$, number of iterations T.
Output: learned transforms $\{\tilde{\mathbf{\Omega}}_l^{(T)}\}$.
for $t = 1, 2, \cdots, T$ **do**
    for $l = 1, 2, \cdots, L$ **do**
        1) Sparse Coding for $\tilde{\mathbf{Z}}_l^{(t)}$ via (4).
        2) Updating $\tilde{\mathbf{\Omega}}_l^{(t)}$ via (8).
    end for
end for

II.B.2. Image Reconstruction Algorithm

The proposed PWLS-MARS algorithm for low-dose CT image reconstruction exploits the learned model. We reconstruct the image by solving the PWLS problem (P1). We propose a block coordinate descent (BCD) algorithm for (P1) that cycles over updating the image $\mathbf{x}$ and each of the sparse coefficient maps $\mathbf{Z}_l$ for $1 \le l \le L$.

(a) Image Update Step for $\mathbf{x}$

First, with the sparse coefficient maps $\{\mathbf{Z}_l\}$ fixed, we optimize for $\mathbf{x}$ in (P1) by optimizing the following subproblem:

$$\min_{\mathbf{x}\geq\mathbf{0}} \frac{1}{2}\|\mathbf{y}-\mathbf{A}\mathbf{x}\|^2_{\mathbf{W}} + \beta \mathsf{S}_2(\mathbf{x}), \tag{9}$$

where $\mathsf{S}_2(\mathbf{x}) \triangleq \sum_{l=1}^{L}\left\{\|\mathbf{\Omega}_l\mathbf{R}_l - \mathbf{Z}_l\|_F^2\right\}$, with $\mathbf{R}_l = \mathbf{\Omega}_{l-1}\mathbf{R}_{l-1} - \mathbf{Z}_{l-1}$, $2 \leq l \leq L$, and $\mathbf{R}_1^j = \mathbf{P}^j\mathbf{x}$. We use the efficient relaxed linearized augmented Lagrangian method[31] (relaxed LALM) to obtain the solution to (9). The algorithmic details are shown in **Algorithm 2**. In each iteration of the relaxed LALM, we update the image T_i times (corresponding to T_i inner loops in **Algorithm 2**). We let matrix $\mathbf{D_A}$ denote a diagonal majorizing matrix of $\mathbf{A}^T\mathbf{W}\mathbf{A}$ and precompute the Hessian matrix of $\mathsf{S}_2(\mathbf{x})$ as $\mathbf{D}_{\mathsf{S}_2}$ in (11) to accelerate the algorithm, and the gradient of $\mathsf{S}_2(\mathbf{x})$ is shown in (10), where $(\mathbf{B}_0^k)^j$ denotes the jth column of matrix $\mathbf{B}_0^k$. We decrease the parameter ρ in Algorithm 2 according to (12)[31], where r denotes the index of inner iterations and the relaxation parameter $\alpha \in [1, 2)$ in (12).

$$\nabla \mathsf{S}_2(\mathbf{x}) = 2\beta \sum_{j=1}^{N_p} (\mathbf{P}^j)^T \left\{ L\mathbf{P}^j\mathbf{x} - \sum_{k=1}^{L}(\mathbf{B}_0^k)^j \right\}, \tag{10}$$

$$\mathbf{D}_{\mathsf{S}_2} \triangleq \nabla^2 \mathsf{S}_2(\mathbf{x}) = 2L\beta \sum_{j=1}^{N_p} (\mathbf{P}^j)^T\mathbf{P}^j, \tag{11}$$

$$\rho_r(\alpha) = \begin{cases} 1, & r = 0, \\ \frac{\pi}{\alpha(r+1)}\sqrt{1-(\frac{\pi}{2\alpha(r+1)})^2}, & \text{otherwise}, \end{cases} \tag{12}$$

(b) Sparse Coding Step for Each $\mathbf{Z}_l$

Similar to the sparse coding step during transform learning, the solution of (P1) with respect to each sparse coefficient map $\mathbf{Z}_l$ is shown in (14), and is the solution of (13).

$$\begin{aligned} &\min_{\mathbf{Z}_l} \sum_{i=l}^{L}\left\{\|\mathbf{\Omega}_i\mathbf{R}_i - \mathbf{Z}_i\|_F^2\right\} + \gamma_l^2\|\mathbf{Z}_l\|_0, \\ &\text{s.t.} \quad \mathbf{R}_i = \mathbf{\Omega}_{i-1}\mathbf{R}_{i-1} - \mathbf{Z}_{i-1}, \quad l \leq i \leq L, \end{aligned} \tag{13}$$

$$\hat{\mathbf{Z}}_l = H_{\gamma_l/\sqrt{L-l+1}}\left\{\mathbf{\Omega}_l\mathbf{R}_l - \frac{1}{L-l+1}\sum_{i=l+1}^{L}\mathbf{B}_l^i\right\}. \tag{14}$$

Algorithm 2 Image Reconstruction Algorithm

Input: initial image $\tilde{\mathbf{x}}^{(0)}$, all-zero initial $\{\tilde{\mathbf{Z}}_l^{(0)}\}$, pre-learned $\{\mathbf{\Omega}_l\}$, thresholds $\{\gamma_l\}$, $\alpha = 1.999$, $\mathbf{D_A}$, $\mathbf{D}_{\mathsf{S}_2}$, number of outer iterations T_O, number of inner iterations T_i.
Output: reconstructed image $\tilde{\mathbf{x}}^{(T_O)}$.
for $t = 0, 1, 2, \cdots, T_O - 1$ **do**
  1) Image Update: With $\{\tilde{\mathbf{Z}}_l^{(t)}\}$ fixed,
  Initialization: $\rho = 1$, $\mathbf{x}^{(0)} = \tilde{\mathbf{x}}^{(t)}$, $\mathbf{g}^{(0)} = \boldsymbol{\zeta}^{(0)} = \mathbf{A}^T\mathbf{W}(\mathbf{A}\mathbf{x}^{(0)} - \mathbf{y})$ and $\mathbf{h}^{(0)} = \mathbf{D_A}\mathbf{x}^{(0)} - \boldsymbol{\zeta}^{(0)}$.
  for $r = 0, 1, 2, \cdots, T_i - 1$, **do**

$$\begin{cases} \mathbf{s}^{(r+1)} = \rho(\mathbf{D_A}\mathbf{x}^{(r)} - \mathbf{h}^{(r)}) + (1-\rho)\mathbf{g}^{(r)} \\ \mathbf{x}^{(r+1)} = [\mathbf{x}^{(r)} - (\rho\mathbf{D_A} + \mathbf{D}_{\mathsf{S}_2})^{-1}(\mathbf{s}^{(r+1)} + \nabla\mathsf{S}_2(\mathbf{x}^{(r)}))]_+ \\ \boldsymbol{\zeta}^{(r+1)} \triangleq \mathbf{A}^T\mathbf{W}(\mathbf{A}\mathbf{x}^{(r+1)} - \mathbf{y}) \\ \mathbf{g}^{(r+1)} = \dfrac{\rho}{\rho+1}(\alpha\boldsymbol{\zeta}^{(r+1)} + (1-\alpha)\mathbf{g}^{(r)}) + \dfrac{1}{\rho+1}\mathbf{g}^{(r)} \\ \mathbf{h}^{(r+1)} = \alpha(\mathbf{D_A}\mathbf{x}^{(r+1)} - \boldsymbol{\zeta}^{(r+1)}) + (1-\alpha)\mathbf{h}^{(r)} \end{cases}$$

   decreasing ρ using (12).
  end for
  $\tilde{\mathbf{x}}^{(t+1)} = \mathbf{x}^{(T_i)}$.
  2) Sparse Coding: with $\tilde{\mathbf{x}}^{(t+1)}$ fixed, for each $1 \le l \le L$, update $\tilde{\mathbf{Z}}_l^{(t+1)}$ sequentially by (14).
end for

III. Experiments

In this section, we evaluate the image reconstruction quality for the proposed PWLS-MARS algorithm and compare it with several conventional or related methods:

- **FBP**: conventional FBP method with a Hanning window.

- **PWLS-EP**[32]: PWLS reconstruction combined with the edge-preserving regularizer $\mathsf{R}(\mathbf{x}) = \sum_{\mathbf{j=1}}^{\mathbf{N_p}} \sum_{\mathbf{k} \in \mathbf{N_j}} \kappa_{\mathbf{j}}\kappa_{\mathbf{k}}\phi(\mathbf{x_j} - \mathbf{x_k})$, where N_j denotes the set of neighborhood pixel indices, and κ_j and κ_k are the parameters that encourage uniform noise[32]. We use $\phi(t) \triangleq \delta^2(\sqrt{1 + \|t/\delta\|^2} - 1)$ as the potential function. The relaxed OS-LALM[31] is the chosen optimizing approach for this PWLS cost function.

To compare the image quality quantitatively, we compute the root mean square error (RMSE) and the structural similarity index measure (SSIM)[4,33]. The RMSE in Hounsfield units (HU) is computed between the ground truth image and reconstructed image as RMSE $= \sqrt{\Sigma_{i \in \mathrm{ROI}}(\hat{x}_i - x_i^*)^2 / N_{\mathrm{ROI}}}$, where $\hat{x}_i$ and x_i^* denote the pixel intensities of the reconstructed and ground truth images, respectively, and N_{ROI} is the number of pixels in the region of interest (ROI). The ROI here was a circular (around center of image) region containing all the phantom tissues. We simulate the low-dose CT measurements using the "Poisson + Gaussian" noisy model[34], i.e., $\hat{\mathbf{y}}_{\mathbf{i}} = \text{Poisson}\{I_0 e^{-[\mathbf{A}\mathbf{x}]_{\mathbf{i}}}\} + \mathcal{N}\{0, \sigma^2\}$, where I_0 is the incident X-ray intensity incorporating X-ray source illumination and the detector gain, and $\sigma^2 = 5^2$ is the variance of electronic noise[34].

We conduct experiments with the XCAT phantom[35] and Mayo Clinic data[36], respectively. Our first experiment uses the XCAT phantom data with a clean ground truth (reference) to demonstrate the performance of the MARS model over other schemes and illustrates the learned multi-layer filters. In our second experiment, we investigate the performance of various methods on the Mayo Clinic data and provide a more detailed comparison between MARS and other methods. Lastly, we analyze the residual maps in the proposed model in different layers to better understand the MARS model.

III.A. Parameter Selection

For each MARS model, multiple parameters are tuned for the learning ($\{\eta_l, 1 \leq l \leq L\}$) and reconstruction ($\beta, \{\gamma_l, 1 \leq l \leq L\}$) stages. Even though the number of parameters here increases the difficulty of adjusting the model for optimal image quality, we can choose the values of the parameters with an empirical approach. The parameters $\{\eta_l\}$ during learning are to achieve a low sparsity of the sparse coefficient maps. Normally, we set $\{\eta_l\}$ to achieve $5 - 10\%$ sparsity for $\mathbf{Z}_l$. One clever method for selecting good sparsity penalty parameters is to set them in decreasing order over layers. This strategy is expected to work because the residual maps in subsequent layers always contain less (or finer) image information than the early layers. A similar approach works for adjusting parameters in the reconstruction stage. In the reconstruction algorithm, we tune the parameters over ranges of values (decreasing over layers for γ_l) to achieve the best reconstruction quality (i.e., RMSE and SSIM).

III.B. Results with the XCAT Phantom

III.B.1. Behavior of the Learned MARS Models

We pre-learn MARS models with different numbers of layers (depths) with 64×64 transforms. The models are learned from 8×8 overlapping patches extracted from five 420×420 XCAT phantom slices. The number of pixels N_p and the number of overall training patches N are about 1.7×10^5 and 8.5×10^5, respectively. The training slices are displayed in the supplement (Fig. S-1). The patch stride is 1×1. We choose 1, 2, 3, 5, and 7 layers, respectively, during training, which corresponds to ST, MARS2, MARS3, MARS5, and MARS7 models. We initialize the MARS learning algorithm with the 2D DCT matrix for the transform in the first layer and identity matrices for transforms in deeper layers. For each model, we ran 1000 to 1500 iterations of the block coordinate descent training algorithm to ensure convergence. We choose $\eta = 75$ for ST, $(\eta_1, \eta_2) = (80, 60)$ for MARS2, $(\eta_1, \eta_2, \eta_3) = (90, 80, 60)$ for MARS3, $(\eta_1, \eta_2, \eta_3, \eta_4, \eta_5) = (120, 120, 120, 110, 110)$ for MARS5, $(\eta_1, \eta_2, \eta_3, \eta_4, \eta_5, \eta_6, \eta_7) = (120, 120, 120, 110, 110, 80, 60)$ for MARS7. Fig. 3 shows some of the learned transforms, with each transform matrix row displayed as a square patch for simplicity. The first layer transform in the models typically displays edge-like and gradient filters that sparsify the image. However, with more layers, finer level features are learned to sparsify transform-domain residuals in deeper layers. Nonetheless, the transforms in quite deep layers could potentially be more easily contaminated with noise in the training data, since the main image features are successively filtered out over layers.

III.B.2. Simulation Framework and Visual Results

We simulate low-dose CT measurements using 840×840 XCAT phantom slices with $\Delta_x = \Delta_y = 0.4883$ mm. The generated sinograms are of size 888×984, obtained with GE 2D LightSpeed fan-beam geometry corresponding to a monoenergetic source with $I_0 = 1 \times 10^4$ incident photons per ray and no scatter. For PWLS-EP, we ran 1000 iterations of the relaxed LALM algorithm with the FBP reconstruction as initialization and regularization parameter $\beta = 2^{16}$. For the MARS model, we used the relaxed LALM algorithm for the image update step with 2 inner iterations. We initialized PWLS-MARS schemes with the PWLS-EP reconstruction and used $T_O = 1500$ outer iterations for ST and all MARS schemes.

We firstly hand-tuned the reconstruction parameters $(\beta, \{\gamma_l, 1 \leq l \leq L\})$ for one test slice and treated this set of parameters as the baseline. Similar to the PWLS-EP algorithm, we could determine the optimal (in terms of optimal RMSE) parameters for other testing slices by tuning the base parameters in a small range. However, we found that the change in reconstruction quality by picking a common set of parameters instead of slice-wise optimized parameters is quite small (only 0.2 HU in RMSE and without the loss of details). Therefore, the same set of parameters (baseline parameters) were used across testing cases and shown to be effective over the cases. In particular, we selected slice 48 of the XCAT phantom as the case for parameter tuning and set the regularization parameters (after tuning over ranges of values) as $(\beta, \gamma) = (2 \times 10^5, 20)$ for ST, $(\beta, \gamma_1, \gamma_2) = (9 \times 10^4, 30, 10)$ for MARS2, $(\beta, \gamma_1, \gamma_2, \gamma_3) = (9 \times 10^4, 25, 15, 10)$ for MARS3, $(\beta, \gamma_1, \gamma_2, \gamma_3, \gamma_4, \gamma_5) = (9 \times 10^4, 25, 15, 10, 5, 1)$ for MARS5, and $(\beta, \gamma_1, \gamma_2, \gamma_3, \gamma_4, \gamma_5, \gamma_6, \gamma_7) = (6 \times 10^4, 30, 25, 20, 15, 10, 5, 1)$ for MARS7, respectively. In Fig. S-2 in the supplement, we give the reconstructions for slice 48 of the XCAT phantom with various methods. Figs. 4 and 5 here show the reconstructions for two independent test cases (slice 20 and 60 of the XCAT phantom). Both of them used the same set of parameters obtained for slice 48. The zoom-in regions give an explicit comparison between the multi-layer sparsifying transform models and other methods such as FBP, PWLS-EP, and PWLS-ST. PWLS-MARS achieves better noise reduction and higher contrast.

III.C. Low-dose Experiments with Mayo Clinic Data

III.C.1. Study of Model Training

First, we study transform training based on Mayo Clinic data. As shown in Fig. 6, seven 512×512 slices obtained at regular dose from three patients are used for transform learning. The number of pixels $N_p \approx 2.6 \times 10^5$. Similar to the phantom experiments, 8×8 overlapping patches are extracted with a 1×1 patch stride. The number of overall training patches N is about 1.8×10^6. We set $\eta = 100$ for ST, $(\eta_1, \eta_2) = (80, 60)$ for MARS2, $(\eta_1, \eta_2, \eta_3) = (60, 60, 40)$ for MARS3, $(\eta_1, \eta_2, \eta_3, \eta_4, \eta_5) = (100, 100, 80, 80, 60)$ for MARS5, $(\eta_1, \eta_2, \eta_3, \eta_4, \eta_5, \eta_6, \eta_7) = (150, 140, 130, 120, 110, 100, 90)$ for MARS7. The iteration number $T = 1000$ in **Algorithm 1.** Fig. 7 illustrates the learned transforms obtained with Mayo Clinic data. Different from the XCAT phantom case, these transforms up to MARS5 display more complex features and structures. The rich features of the MARS models better sparsify the training images over layers compared to the single-layer model (ST).

III.C.2. Simulation Framework, Reconstruction Results, and Comparisons

The synthesized low-dose clinical measurements are simulated from regular-dose images at a resolution of $\Delta_x = \Delta_y = 0.9766$ mm with a fan-beam CT geometry corresponding to a monoenergetic source at incident photon intensity $I_0 = 1 \times 10^4$. The sinograms are of size 736×1152. The width of each detector column is 1.2858 mm, the source to detector distance is 1085.6 mm, and the source to rotation center distance is 595 mm. We reconstruct images of size 512×512 with the pixel size being 0.69 mm $\times$ 0.69 mm.

We conducted experiments on one test slice used for parameter tuning (L067-slice 120) and four independent test slices (L109-slice 90, L192-slice90, L333-slice140, L506-slice 100) of the Mayo Clinic data. For PWLS-EP, we ran 1000 iterations using relaxed OS-LALM and set regularization parameter $\beta = 2^{15.5}$. We used the same $T_O = 1500$ as the phantom experiments for **Algorithm 2**. The process of selecting a general set of reconstruction parameters $(\beta, \{\gamma_l, 1 \leq l \leq L\})$ for the Mayo Clinic test slices is identical to that for the XCAT phantom in Section III.B.2. The selected regularization parameter β and the parameters γ_l that control the sparsity of the coefficient maps are $(\beta, \gamma) = (2.5 \times 10^4, 30)$ for ST, $(\beta,\ \gamma_1,\ \gamma_2) = (1.8 \times 10^4,\ 30,\ 10)$ for MARS2, $(\beta,\ \gamma_1,\ \gamma_2,\ \gamma_3) = (1.8 \times 10^4,\ 30,\ 12,\ 10)$ for MARS3, $(\beta,\ \gamma_1,\ \gamma_2,\ \gamma_3,\ \gamma_4,\ \gamma_5) = (1.6 \times 10^4,\ 30,\ 20,\ 10,\ 7,\ 5)$ for MARS5, and $(\beta,\ \gamma_1,\ \gamma_2,\ \gamma_3,\ \gamma_4,\ \gamma_5,\ \gamma_6,\ \gamma_7) = (3.5 \times 10^4,\ 20,\ 17,\ 14,\ 11,\ 7,\ 4,\ 1)$ for MARS7, respectively.

Figs. 8, 9, 10, and 11 show the reconstructions of the four independent slices using the FBP, PWLS-EP, PWLS-ST, PWLS-MARS2, PWLS-MARS3, PWLS-MARS5, and PWLS-MARS7 schemes, respectively. Additional Mayo Clinic experimental results of the parameter tuning case (Fig. S-3) are shown in the supplementary document. Table 1 lists the RMSE and SSIM values of reconstructions of the four independent test slices, with the best values bolded. Generally, the five and seven layer models provided the best RMSE and SSIM values. They outperform the single-layer model by 1.9 HU in RMSE on average. However, the MARS5 and MARS7 models perform similarly. In order to strengthen the benefits of the multi-layer model, Table 2 lists the RMSE of the reconstructions in four different ROIs (shown in the reference of Fig. 11) with seven methods for slice 100 of patient L506. By observing the reconstructed images, we see that although the ST model achieves a cleaner reconstruction result than FBP and PWLS-EP, it still sacrifices some sharpness of the central region and suffers from loss of details. The deeper models have a somewhat more positive effect in terms of maintaining subtle features, which is clearly more essential to clinical diagnosis. Furthermore, as we will discuss later, after considerable parameter tuning, we found that the information contained in residual maps is gradually decreased with the number of layers, eventually vanishing at some layer, which suggests that very deep unsupervised models might not offer significantly better image quality.

III.C.3. Analysis of Residual Maps

Here, we investigate the residual images over the layers of the MARS7 model. Fig. 12 displays the image reconstructed with MARS7 along with the residual images in different layers. The residual images are generated by applying the restoring operation $(\mathbf{P}^j)^T$ to the corresponding columns of each residual matrix $\mathbf{R}_l, 1 \leq l \leq L$, forming images $\sum_j (\mathbf{P}^j)^T \mathbf{R}_l^j$. Essentially, all the columns of $\mathbf{R}_l$ are transformed into 8×8 patches and accumulated back in the image to form the residual image in the lth layer. We can observe that the residual images in the first three layers contain explicit structural information and we still find some delicate details in the fourth and fifth layers. However, we hardly see any valuable features in the residual images for the following layers, which is consistent with the fact that the transform is overwhelmed by noise in quite deep layers. Therefore, the ceiling for the potential of multi-layer sparsifying transform model may be 5 or 7 layers. The quantitive result also implies the same conclusion.

III.D. Runtimes for MARS

We also discuss the runtimes for the proposed MARS model. Table 3 shows the average runtimes per iteration (MARS schemes were run for the same overall number of iterations) for various MARS models for both the XCAT phantom and Mayo Clinic data experiments. We ran the Matlab code on a machine with two 2.4GHz 14-core Intel Xeon E5-2680 v4 processors.

We find that although training the deep models (which would be done once offline) takes several times as long as the shallow (single layer) model, the cost of the reconstruction/testing step is much more similar between deep and shallow models.

IV. Discussion and Conclusion

In this work, we presented a strategy for unsupervised learning of deep transform models from limited data and with nested network structure, where the input of each layer comprises of the sparsifiable residual map from the preceding layer. The learned Multi-lAyer Residual Sparsifying transform (MARS) model is used to form a data-driven regularizer in model-based image reconstruction and proves effective for low-dose CT image reconstruction. The proposed algorithms for learning MARS models and for image reconstruction use highly efficient updates and are scalable.

We trained models from patches of (regular-dose) slices of the XCAT phantom and Mayo Clinic data and tested the models for reconstructing other slices. The learned multi-layer models contain complex features and structures, which help enhance image reconstruction quality of MARS models over single layer models. Experiments with both simulated data from the XCAT phantom and with the synthesized clinical data reveal that PWLS-MARS provides better reconstruction metrics and image details compared to other methods such as FBP, PWLS-EP, and PWLS-ST. In Figs. 8, 9, 10, and 11, we observed that the reconstruction incorporating deep transform model prior presented more subtle details, especially for the central region, which normally suffers from severe artifacts in low-dose CT reconstruction.

We also investigated the potential limitation in terms of the model depth. By observing Tables 1 and 2, we found deep models such as MARS7 only offer little additional benefit of RMSE and SSIM. Such a phenomenon also appears in other related work[37] in which the author believes that limited training dataset leads to the deterioration of the performance of deep models. In order to seek the underlying reason, we increased the training dataset from 7 slices to 14 slices while the approximate number of patches to be fed into network has been risen to 3 million. Table 4 lists the reconstruction results of slice 100 of patient L506 with respect to training dataset of 7 slices and 14 slices. The tiny improvement leads us to conjecture that the limitation of the deep model may not be due to the small set of training images. Section. III.C.3. provides an alternative explanation. We found that very deep residual layers may not contain much structures, thus resulting in somewhat noisy transforms there, which may offer little additional benefit.

As shown in Section II.B., the block coordinate descent (BCD) method was applied to train a MARS model. Since the problem we address in this work is nonconvex, there might not be a unique minimizer in general. Despite that we use the BCD algorithm to ensure the monotone decrease over iterations of the nonnegative objective like (P0) with a reasonable initialization (i.e., with PWLS-EP). A more thorough analysis of convergence for our scheme is left for future work.

To conclude, we proposed a general framework for multi-layer residual sparsifying transform (MARS) learning, where the transform domain residual maps over several layers are jointly sparsified. Our work then applied learned MARS models to low-dose CT (LDCT) image reconstruction by using a PWLS approach with a learned MARS regularizer. Experimental results illustrate the promising performance of the multi-layer scheme over single-layer learned sparsifying transforms. Learned MARS models also offer image quality improvements over typical nonadaptive methods. Future work will consider other strategies for learning deep sparsifying models by exploiting pooling and other operations. In addition, more studies are required to validate the proposed method's clinical applicability.

V. Acknowledgments

Xikai Yang and Yong Long are supported in part by NSFC (Grant No. 61501292).

The authors thank Dr. Cynthia McCollough, the Mayo Clinic, the American Association of Physicists in Medicine, and the National Institute of Biomedical Imaging and Bioengineering for providing the Mayo Clinic data.

The authors thank Xuehang Zheng, Shanghai Jiaotong University, China for his helpful suggestions on the experiments.

VI. Conflict of Interest

The authors have no conflicts to disclose.

VII. Data Availability

The data that support the findings of this study are openly available in the National Cancer Institute's The Cancer Imaging Archive (TCIA) at https://doi.org/10.7937/9npb-2637, reference number[36].

Appendix I: Solution of the Sparse Coding Problem (2)

First, we can split this objective function and rewrite (2) as follows,

$$\min_{\mathbf{Z}_l} \|\mathbf{Z}_l - \mathbf{\Omega}_l\mathbf{R}_l\|_F^2 + \sum_{i=l+1}^{L} \|\mathbf{Z}_i - \mathbf{\Omega}_i\mathbf{R}_i\|_F^2 + \eta_l\|\mathbf{Z}_l\|_0. \tag{15}$$

Under the condition that $\mathbf{\Omega}_l^T\mathbf{\Omega}_l = \mathbf{I}, \forall l$, the following steps are based on

$$\|\mathbf{\Omega}_l\mathbf{R}_l - \mathbf{Z}_l\|_F^2 = \|\mathbf{\Omega}_l^T\mathbf{\Omega}_l\mathbf{R}_l - \mathbf{\Omega}_l^T\mathbf{Z}_l\|_F^2 = \|\mathbf{R}_l - \mathbf{\Omega}_l^T\mathbf{Z}_l\|_F^2. \tag{16}$$

We use (16) within (15) repetitively, which leads to the equivalent problem shown in (17),

$$\min_{\mathbf{Z}_l} \|\mathbf{Z}_l - \mathbf{\Omega}_l\mathbf{R}_l\|_F^2 + \sum_{i=l+1}^{L} \|\mathbf{Z}_l + \mathbf{B}_l^i - \mathbf{\Omega}_l\mathbf{R}_l\|_F^2 + \eta_l^2\|\mathbf{Z}_l\|_0. \tag{17}$$

Combining all the quadratic terms involving $\mathbf{Z}_l$ leads to the following optimization problem:

$$\min_{\mathbf{Z}_l}(L-l+1) \times \left\|\mathbf{Z}_l - \left(\mathbf{\Omega}_l\mathbf{R}_l - \frac{1}{L-l+1}\sum_{i=l+1}^{L}\mathbf{B}_l^i\right)\right\|_F^2 + \eta_l^2\|\mathbf{Z}_l\|_0. \tag{18}$$

The solution to (18) is similar to ℓ_0 transform sparse coding[30] and is given as follows when $1 \le l \le L-1$

$$\hat{\mathbf{Z}}_l = H_{\eta_l/\sqrt{L-l+1}}\left(\mathbf{\Omega}_l\mathbf{R}_l - \frac{1}{L-l+1}\sum_{i=l+1}^{L}\mathbf{B}_l^i\right) \tag{19}$$

and when $l = L$, it is given as

$$\hat{\mathbf{Z}}_L = H_{\eta_L}(\mathbf{\Omega}_L \mathbf{R}_L) \tag{20}$$

Appendix II: Solution of the Transform Update Problem (5)

Equation (16) also works well for simplifying (5) as follows,

$$\min_{\mathbf{\Omega}_l : \mathbf{\Omega}_l^T \mathbf{\Omega}_l = \mathbf{I}} (L - l + 1) \times \left\| \mathbf{\Omega}_l \mathbf{R}_l - \mathbf{Z}_l - \frac{1}{L - l + 1} \sum_{i=l+1}^{L} \mathbf{B}_l^i \right\|_F^2 . \tag{21}$$

Problem (21) can be equivalently written as

$$\min_{\mathbf{\Omega}_l : \mathbf{\Omega}_l^T \mathbf{\Omega}_l = \mathbf{I}} tr(\mathbf{R}_l \mathbf{R}_l^T) - 2tr\left(\mathbf{\Omega}_l \mathbf{R}_l \left(\mathbf{Z}_l + \frac{1}{L - l + 1} \sum_{i=l+1}^{L} \mathbf{B}_l^i \right)^T \right). \tag{22}$$

Ignoring the constant first term, we get

$$\max_{\mathbf{\Omega}_l : \mathbf{\Omega}_l^T \mathbf{\Omega}_l = \mathbf{I}} tr\left(\mathbf{\Omega}_l \mathbf{R}_l \left(\mathbf{Z}_l + \frac{1}{L - l + 1} \sum_{i=l+1}^{L} \mathbf{B}_l^i \right)^T \right). \tag{23}$$

Subproblem (23) is identical to the corresponding subproblem in single-layer sparsifying transform learning[30]. We denote the full singular value decomposition of the matrix $\mathbf{G}_l$ as $\mathbf{U}_l \mathbf{\Sigma}_l \mathbf{V}_l^T$. The optimal solution to (23) is then given as $\mathbf{V}_l \mathbf{U}_l^T$ (cf.[30]).

References

[1] 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. **35**, 660–663 (2008).

[2] J. Mairal, M. Elad, and G. Sapiro, Sparse Representation for Color Image Restoration, IEEE Trans. Im. Proc. **17**, 53–69 (2008).

[3] M. Elad and M. Aharon, Image denoising via sparse and redundant representations over learned dictionaries, IEEE Trans. Im. Proc. **15**, 3736–3745 (2006).

[4] Y. Zhang, X. Mou, G. Wang, and H. Yu, Tensor-Based Dictionary Learning for Spectral CT Reconstruction, IEEE Transactions on Medical Imaging **36**, 142–154 (2017).

[5] Y. Pati, R. Rezaiifar, and P. Krishnaprasad, Orthogonal Matching Pursuit: recursive function approximation with applications to wavelet decomposition, in *Asilomar Conf. on Signals, Systems and Computers*, pages 40–44 vol.1, 1993.

[6] M. Aharon, M. Elad, and A. Bruckstein, K-SVD: an algorithm for designing overcomplete dictionaries for sparse representation, IEEE Trans. Sig. Proc. **54**, 4311–4322 (2006).

[7] R. Rubinstein, T. Peleg, and M. Elad, Analysis K-SVD: A dictionary-learning algorithm for the analysis sparse model, IEEE Trans. Sig. Proc. **61**, 661–677 (2013).

[8] L. Le Magoarou and R. Gribonval, Chasing butterflies: In search of efficient dictionaries, in *2015 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 3287–3291, 2015.

[9] S. Ravishankar and Y. Bresler, Learning sparsifying transforms, IEEE Trans. Sig. Proc. **61**, 1072–1086 (2013).

[10] S. Ravishankar and Y. Bresler, Data-Driven Learning of a Union of Sparsifying Transforms Model for Blind Compressed Sensing, IEEE Transactions on Computational Imaging **2**, 294–309 (2016).

[11] B. Wen, Y. Li, and Y. Bresler, When sparsity meets low-rankness: Transform learning with non-local low-rank constraint for image restoration, in *2017 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 2297–2301, 2017.

[12] B. Wen, S. Ravishankar, and Y. Bresler, VIDOSAT: High-Dimensional Sparsifying Transform Learning for Online Video Denoising, IEEE Transactions on Image Processing **28**, 1691–1704 (2019).

[13] Y. LeCun, Y. Bengio, and G. Hionton, Deep learning, Nature **521**, 436–444 (2015).

[14] A. Patel, T. Nguyen, and R. Baraniuk, A Probabilistic Framework for Deep Learning, in *Conference and Workshop on Neural Information Processing Systems (NIPS)*, 2016.

[15] V. Papyan, Y. Romano, and M. Elad, Convolutional Neural Networks Analyzed via Convolutional Sparse Coding, Journal of Machine Learning Research **18**, 2887–2938 (2016).

[16] J. Sulam, V. Papyan, Y. Romano, and M. Elad, Multilayer Convolutional Sparse Modeling: Pursuit and Dictionary Learning, IEEE Transactions on Signal Processing **66**, 4090–4104 (2018).

[17] S. Ravishankar and B. Wohlberg, Learning multi-layer transform models, in *Allerton Conf. on Comm., Control, and Computing*, pages 160–165, 2018.

[18] L. A. Feldkamp, L. C. Davis, and J. W. Kress, Practical cone beam algorithm, J. Opt. Soc. Am. A **1**, 612–619 (1984).

[19] I. A. Elbakri and J. A. Fessler, Statistical image reconstruction for polyenergetic X-ray computed tomography, IEEE Trans. Med. Imag. **21**, 89–99 (2002).

[20] K. Sauer and C. Bouman, A local update strategy for iterative reconstruction from projections, IEEE Trans. Sig. Proc. **41**, 534–548 (1993).

[21] J.-B. Thibault, C. A. Bouman, K. D. Sauer, and J. Hsieh, A recursive filter for noise reduction in statistical iterative tomographic imaging, in *Proc. SPIE*, volume 6065, pages 60650X–1–60650X–10, 2006.

[22] 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. **34**, 4526–4544 (2007).

[23] L. Pfister and Y. Bresler, Model-based iterative tomographic reconstruction with adaptive sparsifying transforms, in *Proc. SPIE*, volume 9020, pages 90200H–1–90200H–11, 2014.

[24] L. Pfister and Y. Bresler, Tomographic reconstruction with adaptive sparsifying transforms, in *Proc. IEEE Conf. Acoust. Speech Sig. Proc.*, pages 6914–6918, 2014.

[25] L. Pfister and Y. Bresler, Adaptive sparsifying transforms for iterative tomographic reconstruction, in *Proc. 3rd Intl. Mtg. on image formation in X-ray CT*, pages 107–110, 2014.

[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. **37**, 1498–1510 (2018).

[27] I. Y. Chun, X. Zheng, Y. Long, and J. A. Fessler, Efficient sparse-view X-ray CT reconstruction using ℓ_1 regularization with learned sparsifying transform, in *Proc. Intl. Mtg. on Fully 3D Image Recon. in Rad. and Nuc. Med*, pages 115–119, 2017.

[28] S. Ye, S. Ravishankar, Y. Long, and J. A. Fessler, SPULTRA: Low-Dose CT Image Reconstruction With Joint Statistical and Learned Image Models, IEEE Transactions on Medical Imaging **39**, 729–741 (2020).

[29] W. Zhou, J.-F. Cai, and H. Gao, Adaptive tight frame based medical image reconstruction: a proof-of-concept study for computed tomography, Inverse Prob. **29**, 125006 (2013).

[30] S. Ravishankar and Y. Bresler, ℓ_0 Sparsifying Transform Learning With Efficient Optimal Updates and Convergence Guarantees, IEEE Trans. Sig. Proc. **63**, 2389–2404 (2015).

[31] H. Nien and J. A. Fessler, Relaxed linearized algorithms for faster X-ray CT image reconstruction, IEEE Trans. Med. Imag. **35**, 1090–1098 (2016).

[32] J. H. Cho and J. A. Fessler, Regularization designs for uniform spatial resolution and noise properties in statistical image reconstruction for 3D X-ray CT, IEEE Trans. Med. Imag. **34**, 678–689 (2015).

[33] Q. Xu, H. Yu, X. Mou, L. Zhang, J. Hsieh, and G. Wang, Low-Dose X-ray CT Reconstruction via Dictionary Learning, IEEE Transactions on Medical Imaging **31**, 1682–1697 (2012).

[34] Q. Ding, Y. Long, X. Zhang, and J. A. Fessler, Modeling mixed Poisson-Gaussian noise in statistical image reconstruction for X-ray CT, in *Proc. 4th Intl. Mtg. on image formation in X-ray CT*, pages 399–402, 2016.

[35] W. P. Segars, M. Mahesh, T. J. Beck, E. C. Frey, and B. M. W. Tsui, Realistic CT simulation using the 4D XCAT phantom, Medical Physics **35**, 3800–3808 (2008).

[36] C. McCollough, TU-FG-207A-04: Overview of the Low Dose CT Grand Challenge., Med. Phys. **43**, 3759–60 (2016).

[37] V. Singhal, J. Maggu, and A. Majumdar, Simultaneous Detection of Multiple Appliances From Smart-Meter Measurements via Multi-Label Consistent Deep Dictionary Learning and Deep Transform Learning, IEEE Transactions on Smart Grid **10**, 2969–2978 (2019).

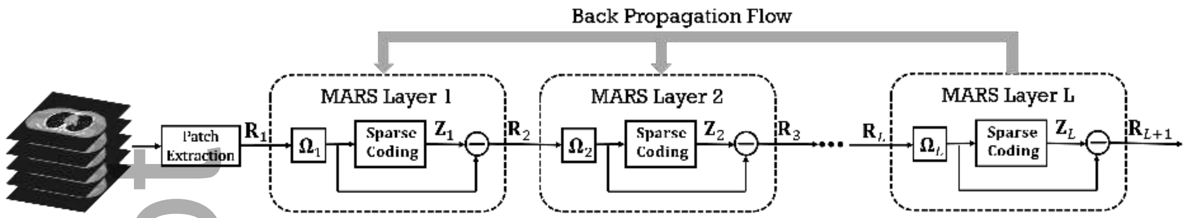


Figure 1: MARS model with L layers or modules. $\mathbf{\Omega}_l$ denotes the transform in the lth layer, which enables sparsifying the residual map arising from the $(l-1)$th module.

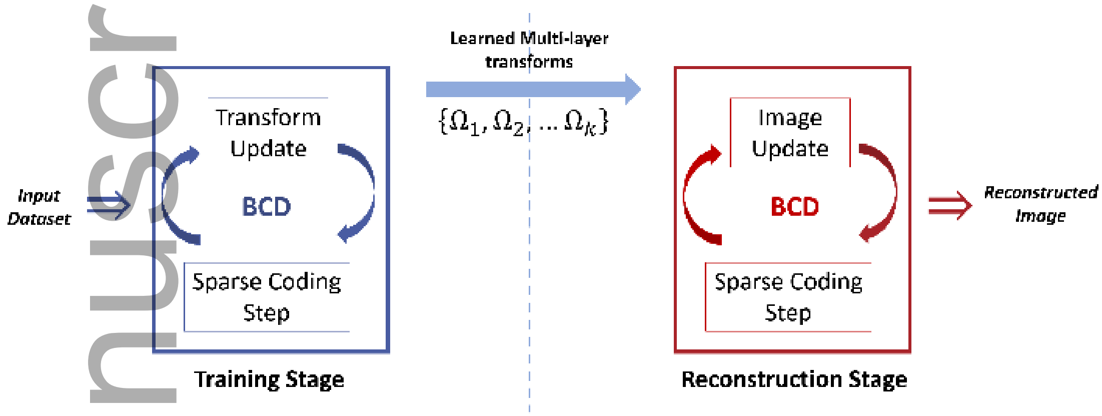


Figure 2: Overview of algorithm scheme. Our approach involves a training stage and a reconstruction stage with block coordinate descent (BCD) algorithms being used in both stages.

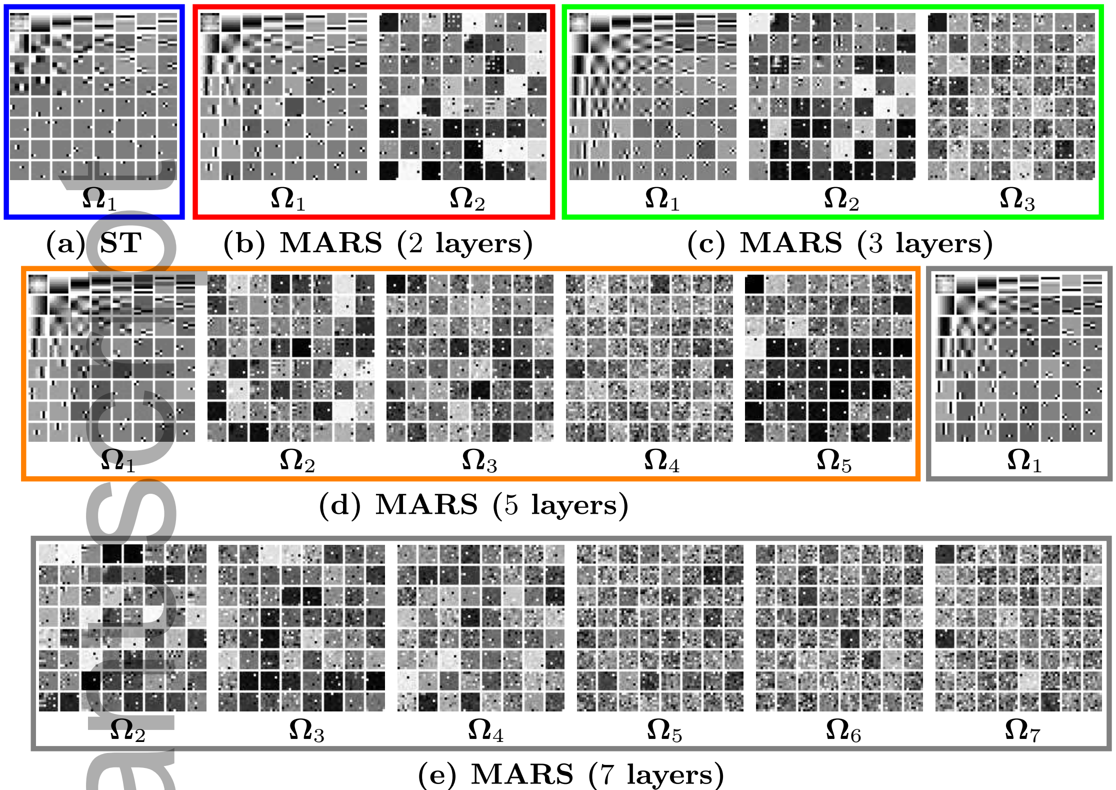


Figure 3: Transforms learned from the XCAT phantom. Transform rows are shown as 8×8 patches. Beyond the first layer, the rows of the transforms sparsify across the residual channels (1D filters).

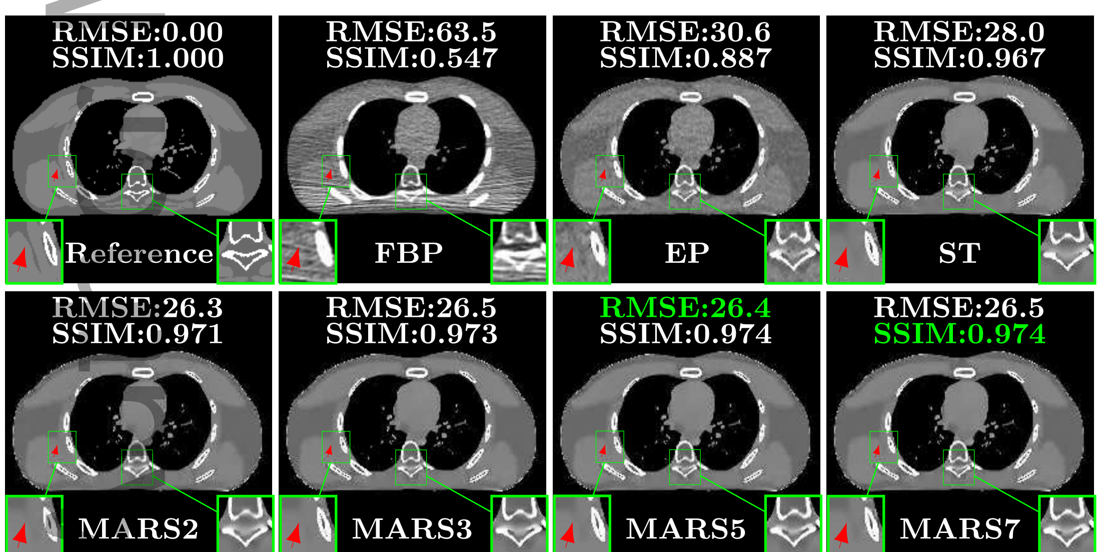


Figure 4: Comparison of reconstructions of slice 20 of the XCAT phantom with FBP, PWLS-EP, PWLS-ST, PWLS-MARS2, PWLS-MARS3, PWLS-MARS5, and PWLS-MARS7, respectively, at incident photon intensity $I_0 = 1 \times 10^4$. The display window is [800, 1200] HU.

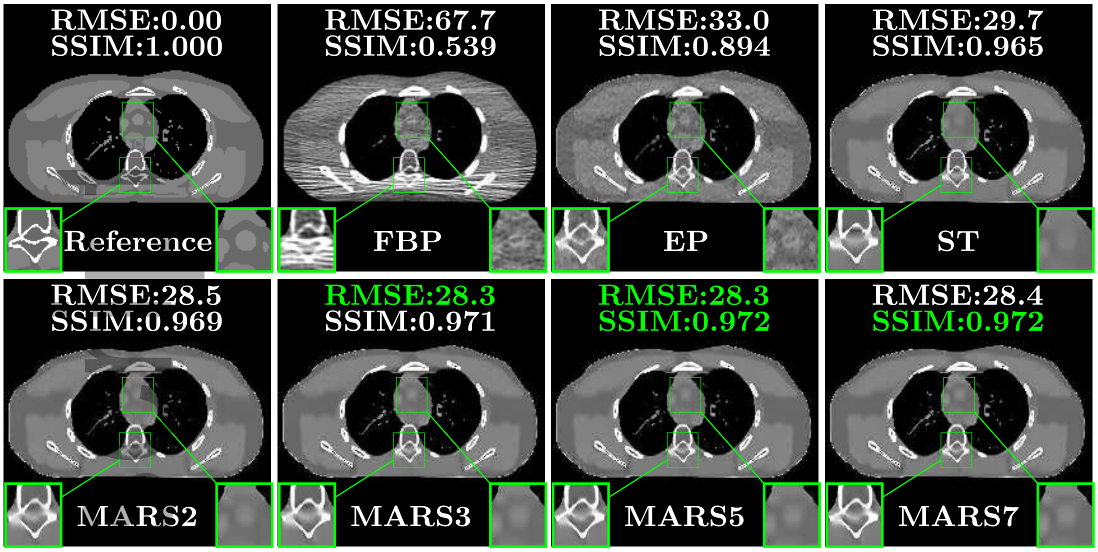


Figure 5: Comparison of reconstructions of slice 60 of the XCAT phantom with FBP, PWLS-EP, PWLS-ST, PWLS-MARS2, PWLS-MARS3, PWLS-MARS5, and PWLS-MARS7, respectively, at incident photon intensity $I_0 = 1 \times 10^4$. The display window is [800, 1200] HU.

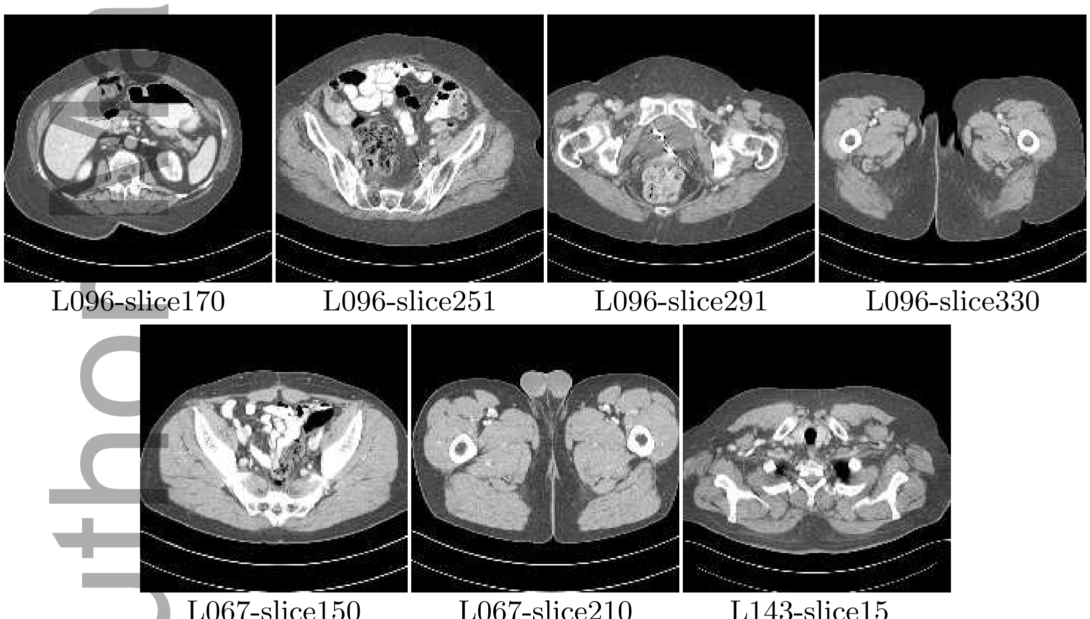


Figure 6: Seven regular-dose slices for training the MARS model. The first row displays four slices of patient L096 and the second row shows three training slices from patients L067 and L143, respectively.

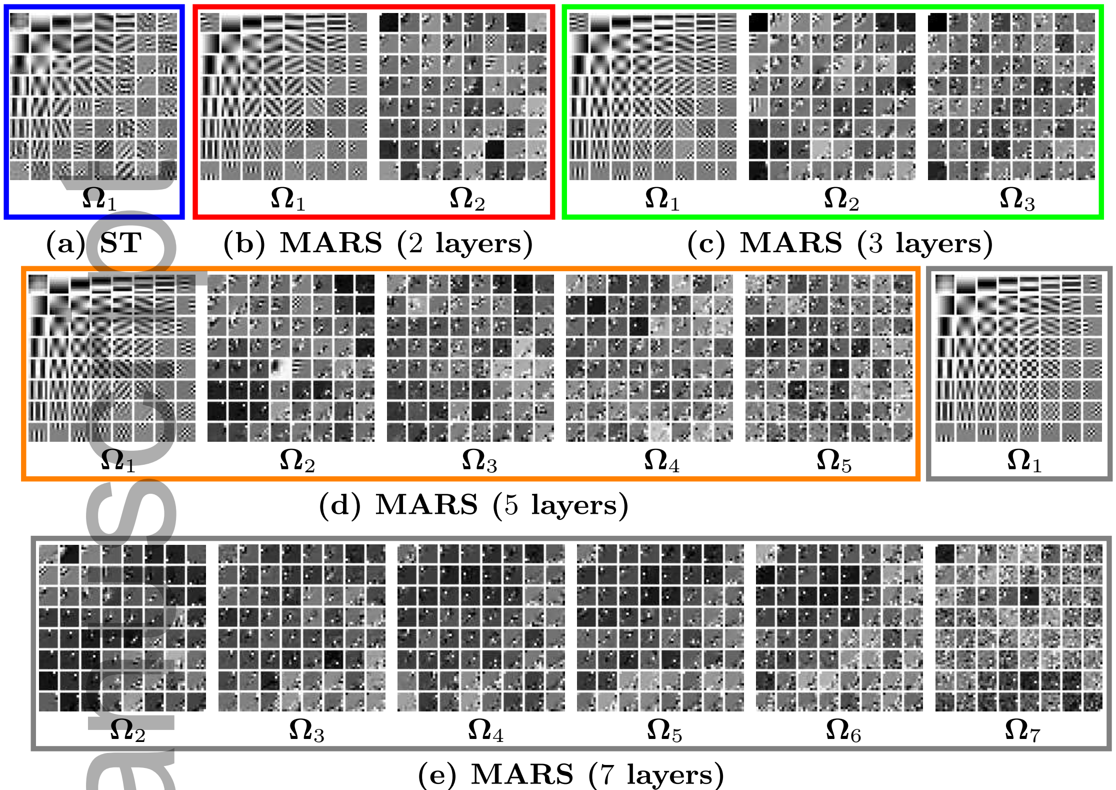


Figure 7: Transforms learned from Mayo Clinic data. Beyond the first layer, the rows of the transforms are shown as (square) 2D patches and sparsify transform-domain residuals.

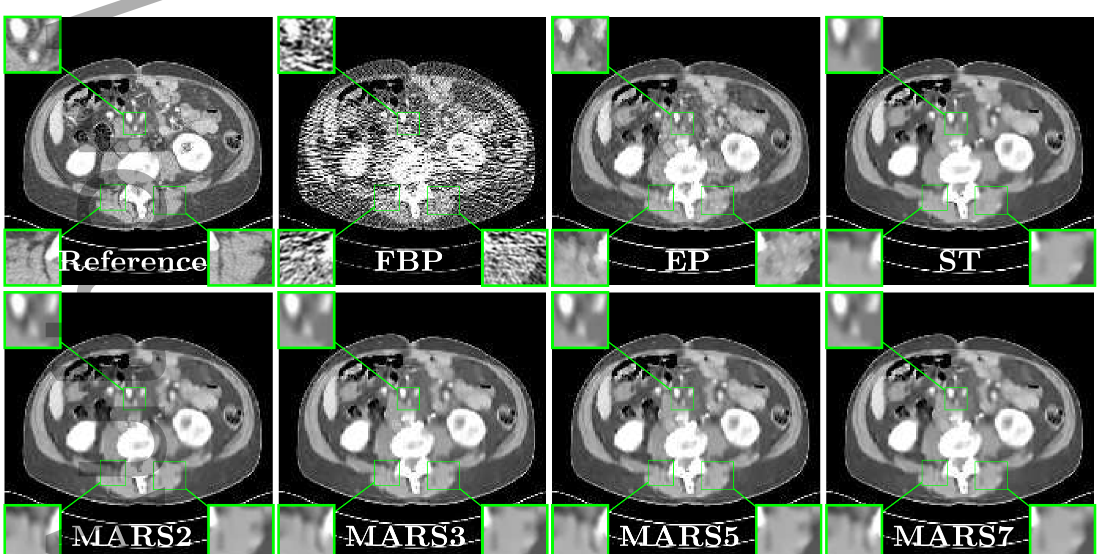


Figure 8: Reconstructions of slice 90 of patient L109 at incident photon intensity $I_0 = 1 \times 10^4$. The first row shows the reference image and reconstructions with FBP, PWLS-EP, and PWLS-ST, respectively, and the second row shows the results with MARS models with 2, 3, 5, and 7 layers, respectively. The display window is [800, 1200] HU.

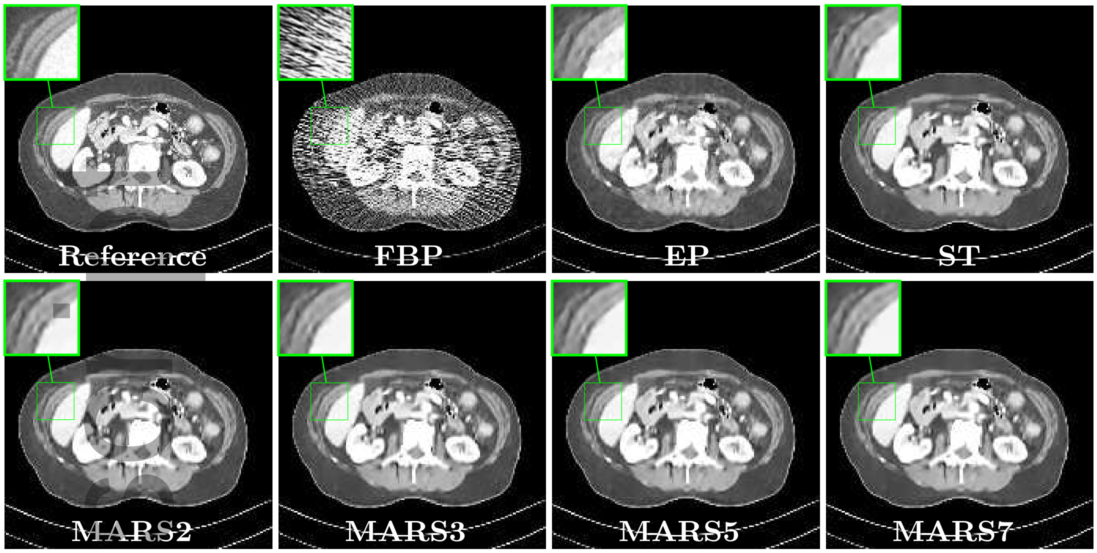


Figure 9: Reconstructions of slice 90 of patient L192 at incident photon intensity $I_0 = 1\times10^4$. The first row shows the reference image and reconstructions with FBP, PWLS-EP, and PWLS-ST, respectively, and the second row shows the results with MARS models with 2, 3, 5, and 7 layers, respectively. The display window is [800, 1200] HU.

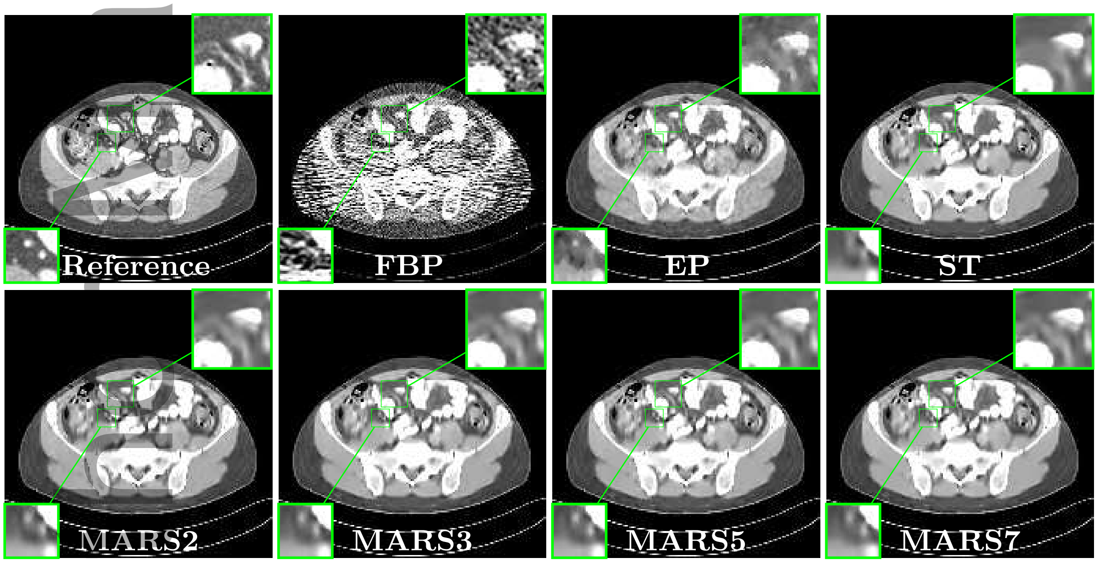


Figure 10: Reconstructions of slice 140 of patient L333 at incident photon intensity $I_0 = 1 \times 10^4$. The first row shows the reference image and reconstructions with FBP, PWLS-EP, and PWLS-ST, respectively, and the second row shows the results with MARS models with 2, 3, 5, and 7 layers, respectively. The display window is [800, 1200] HU.

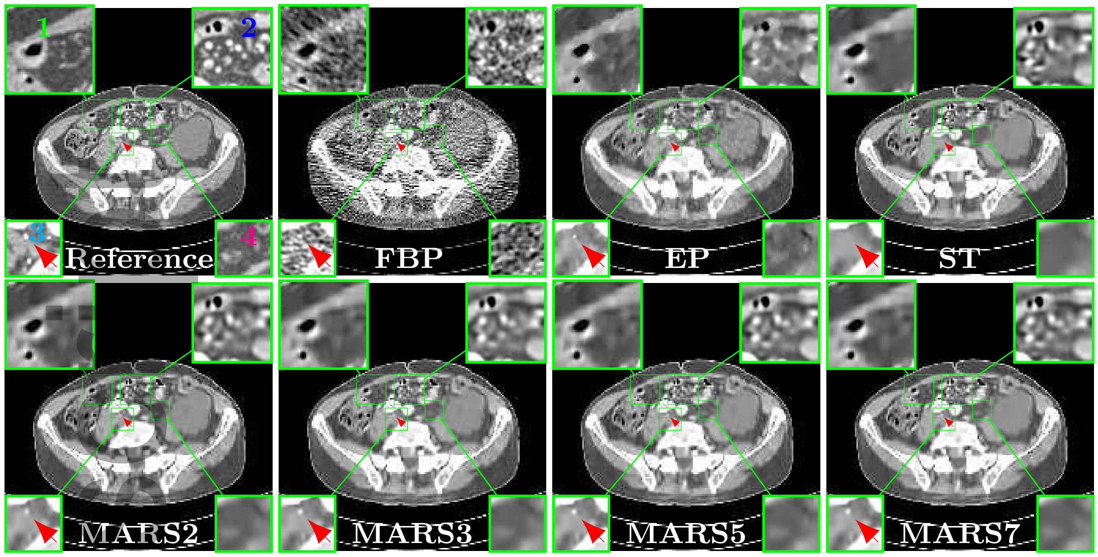


Figure 11: Reconstructions of slice 100 of patient L506 at incident photon intensity $I_0 = 1 \times 10^4$. The first row shows the reference image and reconstructions with FBP, PWLS-EP, and PWLS-ST, respectively, and the second row shows the results with MARS models with 2, 3, 5, and 7 layers, respectively. The display window is [800, 1200] HU.

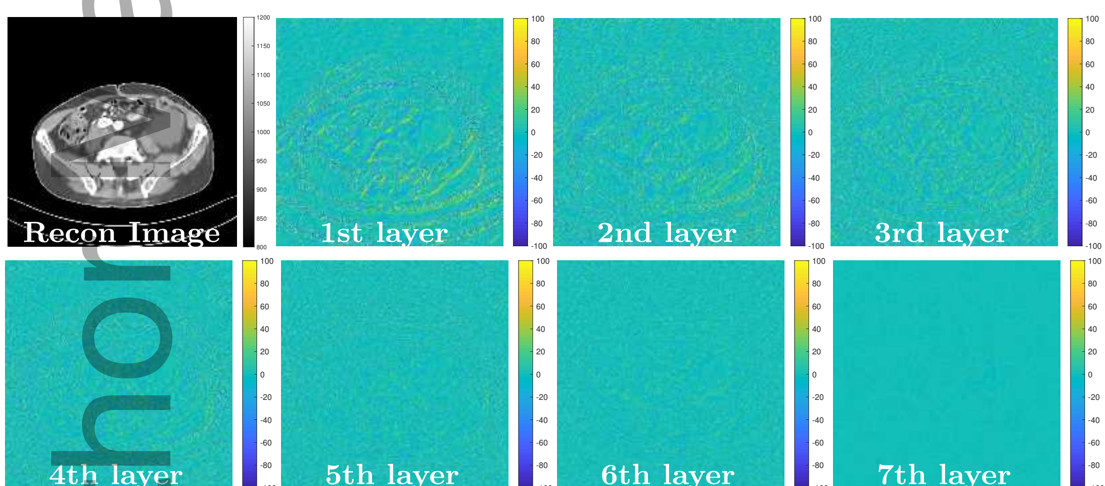


Figure 12: Reconstruction and transform-domain residual images for slice 100 of patient L506. The leftmost image on the first row is the reconstruction with PWLS-MARS7, while the other images are the residual maps in different layers. The display windows are [800, 1200] HU and [-100, 100] HU, respectively, for the reconstruction and the residual image, respectively.

Table 1: RMSE in HU (first row) and SSIM (second row) of reconstructions with FBP, PWLS-EP, PWLS-ST, PWLS-MARS2, PWLS-MARS3, PWLS-MARS5, and PWLS-MARS7, for four slices of the Mayo Clinic data at incident photon intensity $I_0 = 1 \times 10^4$.

	FBP	EP	PWLS-ST	PWLS-MARS2	PWLS-MARS3	PWLS-MARS5	PWLS-MARS7
L109 slice90	107.1	33.5	29.0	28.1	27.8	**27.6**	28.1
	0.343	0.734	0.716	0.727	0.731	0.744	**0.753**
L192 slice90	93.7	31.5	26.3	25.3	24.9	**24.6**	24.9
	0.350	0.747	0.737	0.744	0.750	0.765	**0.781**
L333 slice140	113.1	36.3	29.7	28.5	28.3	**28.1**	28.4
	0.358	0.758	0.739	0.744	0.750	0.766	**0.786**
L506 slice 100	65.3	34.3	27.5	26.2	25.6	**25.3**	25.7
	0.461	0.778	0.760	0.766	0.773	0.790	**0.809**

Table 2: RMSE (HU) in four ROIs of reconstructions with FBP, PWLS-EP, PWLS-ST, PWLS-MARS2, PWLS-MARS3, PWLS-MARS5, and PWLS-MARS7, for slice 100 of patient L506 of the Mayo Clinic data at incident photon intensity $I_0 = 1 \times 10^4$.

	FBP	EP	PWLS-ST	PWLS-MARS2	PWLS-MARS3	PWLS-MARS5	PWLS-MARS7
ROI-1	1.05	0.71	0.68	0.62	0.60	**0.59**	0.59
ROI-2	0.90	0.78	0.69	0.63	0.62	**0.61**	0.63
ROI-3	2.17	1.88	1.75	1.57	1.53	**1.51**	1.55
ROI-4	1.91	0.96	1.03	0.91	0.90	**0.89**	0.91

Table 3: Average runtime per iteration of various MARS models with both XCAT phantom and Mayo Clinic data experiments. Each number displayed in this table is in seconds.

		PWLS-ST	PWLS-MARS2	PWLS-MARS3	PWLS-MARS5	PWLS-MARS7
XCAT phantom	Training	**0.8**	1.4	3.5	4.7	7.8
	Testing	**2.9**	3.2	3.6	4.4	5.1
Mayo Clinic data	Training	**1.5**	2.8	7.4	9.3	15.2
	Testing	**3.1**	3.4	4.1	5.0	5.8

Table 4: Comparison of reconstruction of slice 100 of patient L506 between training dataset of 7 slices and 14 slices respectively.

		PWLS-ST	PWLS-MARS2	PWLS-MARS3	PWLS-MARS5	PWLS-MARS7
dataset of 7 slices	RMSE	27.5	26.2	25.6	**25.3**	25.7
	SSIM	0.760	0.766	0.773	0.790	**0.809**
dataset of 14 slices	RMSE	27.4	26.2	25.6	**25.4**	25.6
	SSIM	0.759	0.766	0.773	0.790	**0.810**

List of Figures:

reconstructions with FBP, PWLS-EP, and PWLS-ST, respectively, and the second row shows the results with MARS models with 2, 3, 5, and 7 layers, respectively. The display window is [800, 1200] HU.

- Figure 10: Reconstructions of slice 140 of patient L333 at incident photon intensity $I_0 = 1 \times 10^4$. The first row shows the reference image and reconstructions with FBP, PWLS-EP, and PWLS-ST, respectively, and the second row shows the results with MARS models with 2, 3, 5, and 7 layers, respectively. The display window is [800, 1200] HU.
- Figure 11: Reconstructions of slice 100 of patient L506 at incident photon intensity $I_0 = 1 \times 10^4$. The first row shows the reference image and reconstructions with FBP, PWLS-EP, and PWLS-ST, respectively, and the second row shows the results with MARS models with 2, 3, 5, and 7 layers, respectively. The display window is [800, 1200] HU.
- Figure 12: Reconstruction and transform-domain residual images for slice 100 of patient L506. The leftmost image on the first row is the reconstruction with PWLS-MARS7, while the other images are the residual maps in different layers. The display windows are [800, 1200] HU and [-100, 100] HU, respectively, for the reconstruction and the residual image, respectively.
- Figure S-1: Five reference slices for training the MARS model. The slice numbers correspond to the location in the volume.
- Figure S-2: Comparison of reconstructions of slice 48 of the XCAT phantom with FBP, PWLS-EP, PWLS-ST, PWLS-MARS2, PWLS-MARS3, PWLS-MARS5, and PWLS-MARS7, respectively, at incident photon intensity $I_0 = 1 \times 10^4$. The display window is [800, 1200] HU.
- Figure S-3: Reconstructions of slice 120 of patient L067 at incident photon intensity $I_0 = 1 \times 10^4$. The first row shows the reference image and reconstructions with FBP, PWLS-EP, and PWLS-ST, respectively, and the second row shows the results with MARS models with 2, 3, 5, and 7 layers, respectively. The display window is [800, 1200] HU.

Table 1: RMSE in HU (first row) and SSIM (second row) of reconstructions with FBP, PWLS-EP, PWLS-ST, PWLS-MARS2, PWLS-MARS3, PWLS-MARS5, and PWLS-MARS7, for four slices of the Mayo Clinic data at incident photon intensity $I_0 = 1 \times 10^4$.

	FBP	EP	PWLS-ST	PWLS-MARS2	PWLS-MARS3	PWLS-MARS5	PWLS-MARS7
L109 Slice90	107.1	33.5	29.0	28.1	27.8	**27.6**	28.1
	0.343	0.734	0.716	0.727	0.731	0.744	**0.753**
L192 Slice90	93.7	31.5	26.3	25.3	24.9	**24.6**	24.9
	0.350	0.747	0.737	0.744	0.750	0.765	**0.781**
L333 Slice140	113.1	36.3	29.7	28.5	28.3	**28.1**	28.4
	0.358	0.758	0.739	0.744	0.750	0.766	**0.786**
L506 Slice100	65.3	34.3	27.5	26.2	25.6	**25.3**	25.7
	0.461	0.778	0.760	0.766	0.773	0.790	**0.809**

Table 2: RMSE (HU) in four ROIs of reconstructions with FBP, PWLS-EP, PWLS-ST, PWLS-MARS2, PWLS-MARS3, PWLS-MARS5, and PWLS-MARS7, for slice 100 of patient L506 of the Mayo Clinic data at incident photon intensity $I_0 = 1 \times 10^4$.

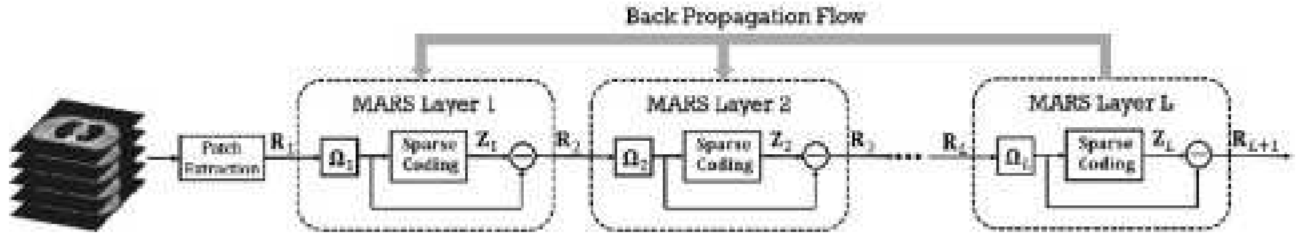
	FBP	EP	PWLS-ST	PWLS-MARS2	PWLS-MARS3	PWLS-MARS5	PWLS-MARS7
ROI-1	1.05	0.71	0.68	0.62	0.60	**0.59**	0.59
ROI-2	0.90	0.78	0.69	0.63	0.62	**0.61**	0.63
ROI-3	2.17	1.88	1.75	1.57	1.53	**1.51**	1.55
ROI-4	1.91	0.96	1.03	0.91	0.90	**0.89**	0.91

Table 3: Average runtime per iteration of various MARS models with both XCAT phantom and Mayo Clinic data experiments. Each number displayed in this table is in seconds.

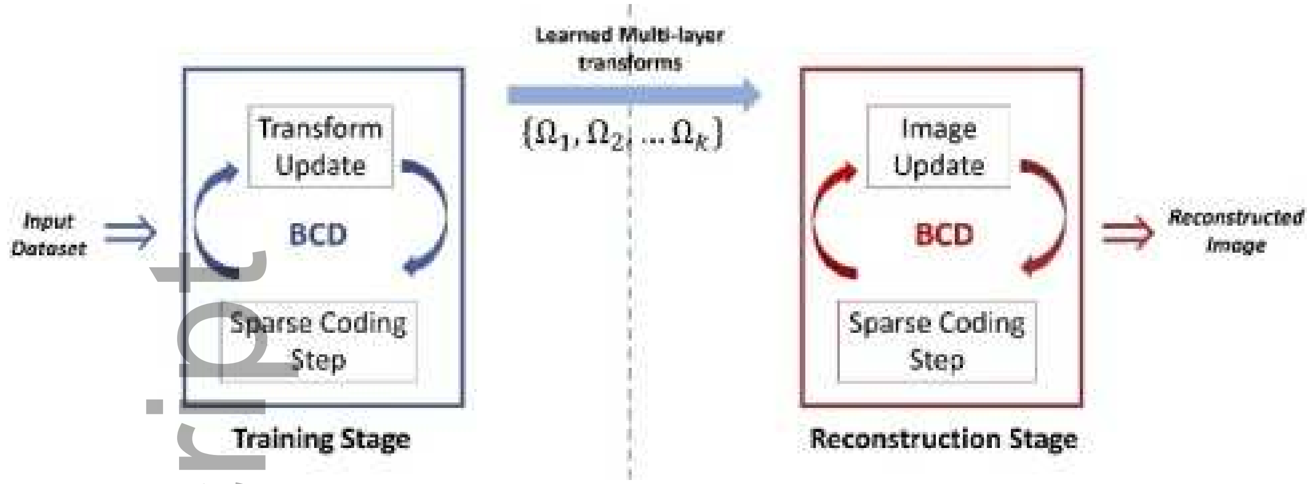
		PWLS-ST	PWLS-MARS2	PWLS-MARS3	PWLS-MARS5	PWLS-MARS7
XCAT phantom	Training	**0.8**	1.4	3.5	4.7	7.8
	Testing	**2.9**	3.2	3.6	4.4	5.1
Mayo Clinic data	Training	**1.5**	2.8	7.4	9.3	15.2
	Testing	**3.1**	3.4	4.1	5.0	5.8

Table 4: Comparison of reconstruction of slice 100 of patient L506 between training dataset of 7 slices and 14 slices respectively.

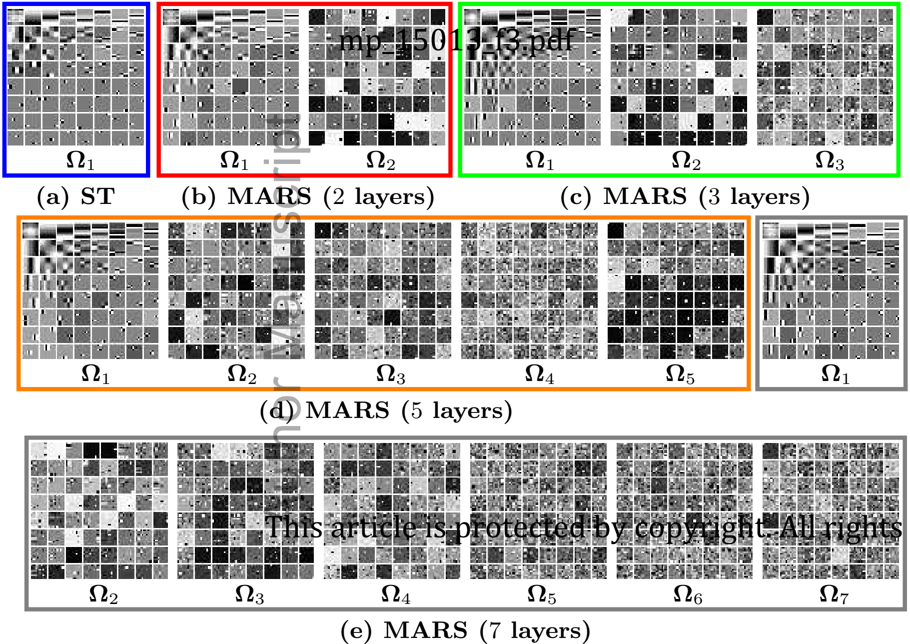
		PWLS-ST	PWLS-MARS2	PWLS-MARS3	PWLS-MARS5	PWLS-MARS7
dataset of 7 slices	RMSE	27.5	26.2	25.6	**25.3**	25.7
	SSIM	0.760	0.766	0.773	0.790	**0.809**
dataset of 14 slice2	RMSE	27.4	26.2	25.6	**25.4**	25.6

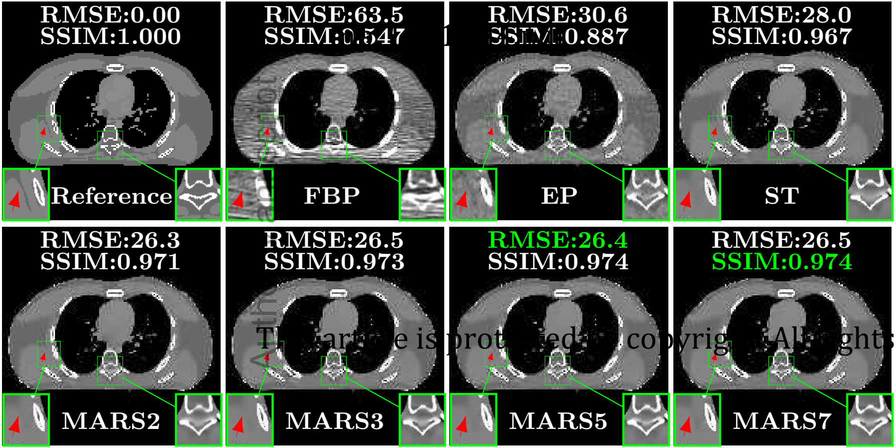


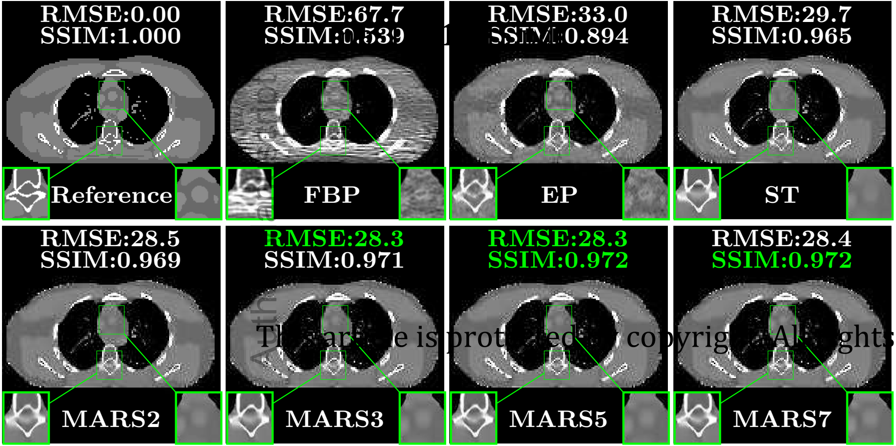
mp_15013_f1.jpg

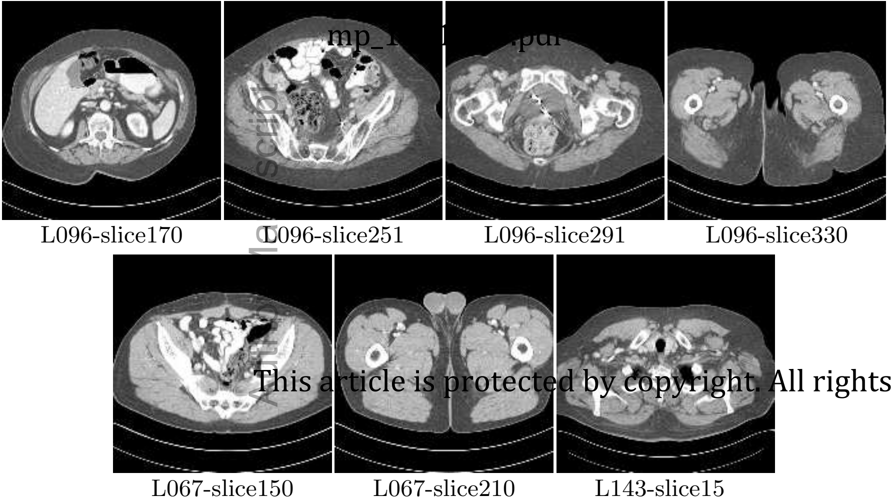


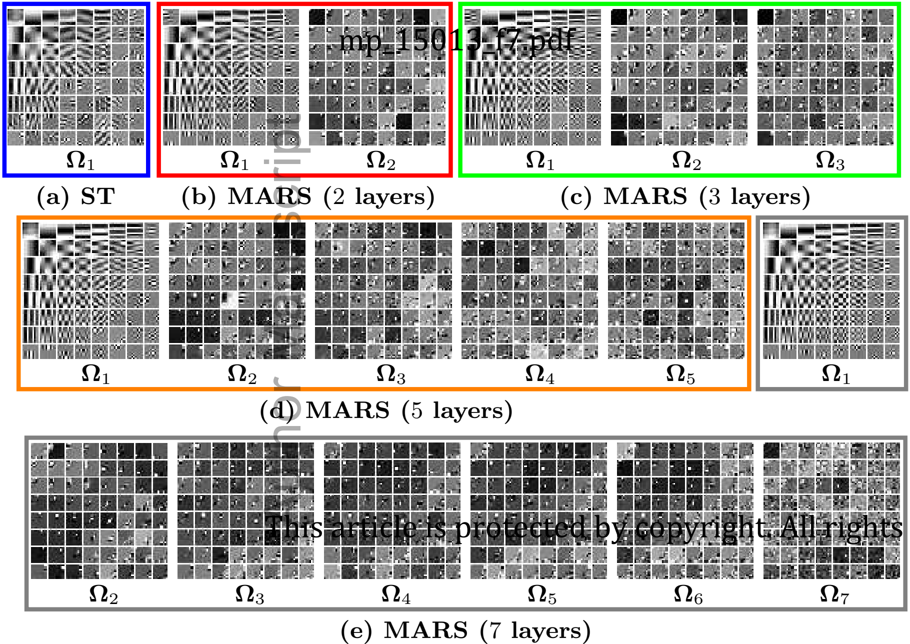
mp_15013_f2.jpg

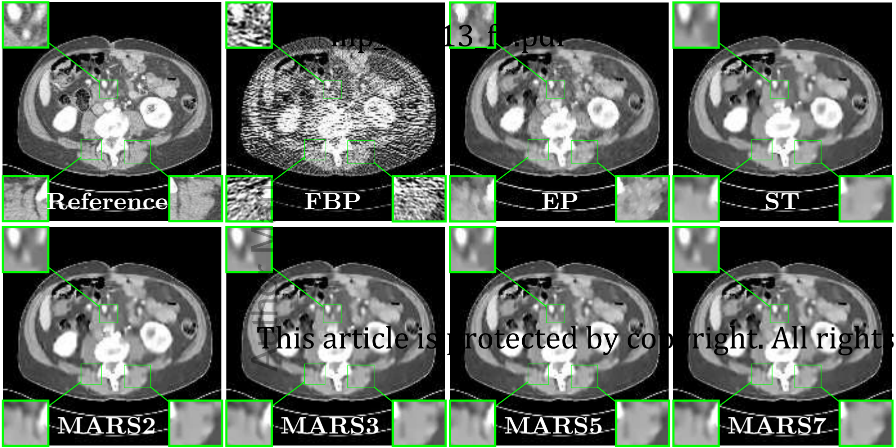

Ω_1
(a) ST
Ω_1 Ω_2
(b) MARS (2 layers)
Ω_1 Ω_2 Ω_3
(c) MARS (3 layers)
Ω_1 Ω_2 Ω_3 Ω_4 Ω_5
(d) MARS (5 layers)
Ω_1 Ω_2 Ω_3 Ω_4 Ω_5 Ω_6 Ω_7
(e) MARS (7 layers)


RMSE:0.00
SSIM:1.000
Reference
RMSE:63.5
SSIM:0.547
FBP
RMSE:30.6
SSIM:0.887
EP
RMSE:28.0
SSIM:0.967
ST
RMSE:26.3
SSIM:0.971
MARS2
RMSE:26.5
SSIM:0.973
MARS3
RMSE:26.4
SSIM:0.974
MARS5
RMSE:26.5
SSIM:0.974
MARS7

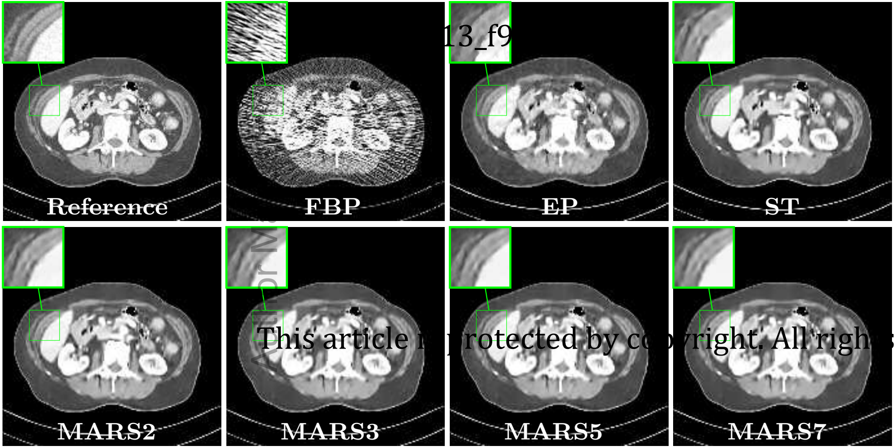

RMSE:0.00
SSIM:1.000
Reference
RMSE:67.7
SSIM:0.539
FBP
RMSE:33.0
SSIM:0.894
EP
RMSE:29.7
SSIM:0.965
ST
RMSE:28.5
SSIM:0.969
MARS2
RMSE:28.3
SSIM:0.971
MARS3
RMSE:28.3
SSIM:0.972
MARS5
RMSE:28.4
SSIM:0.972
MARS7

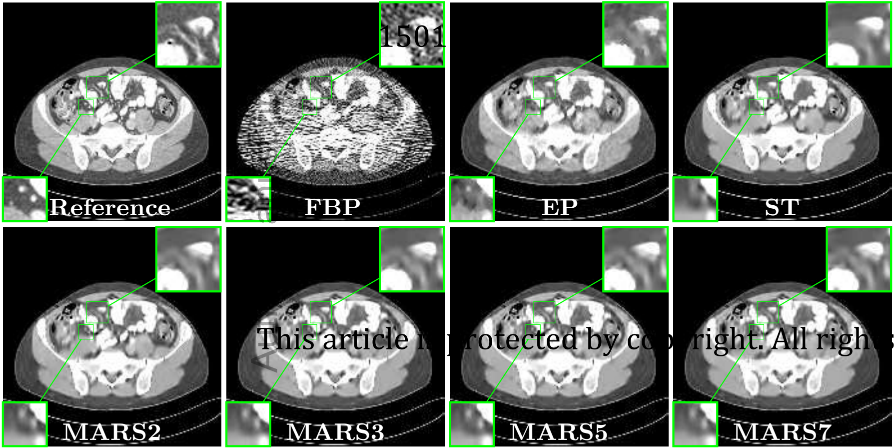

L096-slice170
L096-slice251
L096-slice291
L096-slice330
L067-slice150
L067-slice210
L143-slice15

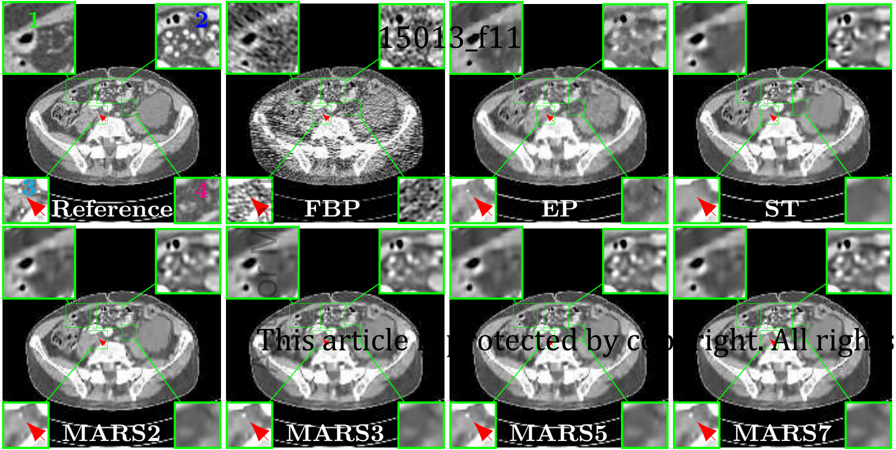

Ω_1
(a) ST
Ω_1 Ω_2
(b) MARS (2 layers)
Ω_1 Ω_2 Ω_3
(c) MARS (3 layers)
Ω_1 Ω_2 Ω_3 Ω_4 Ω_5
(d) MARS (5 layers)
Ω_1 Ω_2 Ω_3 Ω_4 Ω_5 Ω_6 Ω_7
(e) MARS (7 layers)

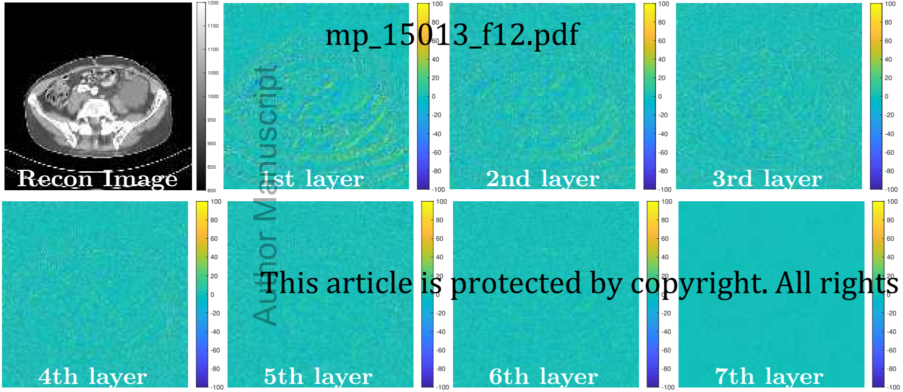

Reference
FBP
EP
ST
MARS2
MARS3
MARS5
MARS7

13_f9




Reference
FBP
EP
ST
MARS2
MARS3
MARS5
MARS7


1
2
3
4
Reference
FBP
EP
ST
MARS2
MARS3
MARS5
MARS7


Recon Image
1st layer
2nd layer
3rd layer
4th layer
5th layer
6th layer
7th layer