SUPER Learning: A Supervised-Unsupervised Framework for Low-Dose CT Image Reconstruction

Zhipeng Li[1]   Siqi Ye[1]   Yong Long[1*]   Saiprasad Ravishankar[2]

[1]University of Michigan - Shanghai Jiao Tong University Joint Institute,
Shanghai Jiao Tong University[†], Shanghai, China

[2]Departments of Computational Mathematics, Science and Engineering,
and Biomedical Engineering, Michigan State University, East Lansing, MI, USA

{zhipengli, yesiqi, yong.long}@sjtu.edu.cn, ravisha3@msu.edu

Abstract

Recent years have witnessed growing interest in machine learning-based models and techniques for low-dose X-ray CT (LDCT) imaging tasks. The methods can typically be categorized into supervised learning methods and unsupervised or model-based learning methods. Supervised learning methods have recently shown success in image restoration tasks. However, they often rely on large training sets. Model-based learning methods such as dictionary or transform learning do not require large or paired training sets and often have good generalization properties, since they learn general properties of CT image sets. Recent works have shown the promising reconstruction performance of methods such as PWLS-ULTRA that rely on clustering the underlying (reconstructed) image patches into a learned union of transforms. In this paper, we propose a new Supervised-UnsuPERvised (SUPER) reconstruction framework for LDCT image reconstruction that combines the benefits of supervised learning methods and (unsupervised) transform learning-based methods such as PWLS-ULTRA that involve highly image-adaptive clustering. The SUPER model consists of several layers, each of which includes a deep network learned in a supervised manner and an unsupervised iterative method that involves image-adaptive components. The SUPER reconstruction algorithms are learned in a greedy manner from training data. The proposed SUPER learning methods dramatically outperform both the constituent supervised learning-based networks and iterative algorithms for LDCT, and use much fewer iterations in the iterative reconstruction modules.

1. Introduction

X-ray computed tomography (CT) is a popular imaging modality in many clinical and industrial applications. There

---

has been particular interest in CT imaging with low X-ray dose levels that would reduce the potential risks to patients from radiation. However, image reconstruction at low X-ray dose levels is challenging. Conventional X-ray CT image reconstruction methods include analytical methods, and model-based iterative reconstruction (MBIR) methods. A classical analytical method is the filtered back-projection (FBP) method [1]. That can be degraded excessively by noise and streak artifacts in low-dose situations [2, 3].

MBIR methods incorporate the system physics, statistical model of measurements, and typically certain simple prior information of the unknown object [4]. A typical method of this kind is the penalized weighted-least squares (PWLS) method, for which the cost function includes a weighted quadratic data-fidelity term that models the measurement statistics, and a penalty term called a regularizer that models the prior information [5–7]. For PWLS, various optimization approaches and regularization designs have been exploited with efficiency and convergence guarantees.

Adopting appropriate prior knowledge of images for MBIR approaches is also important to improve CT reconstruction. More recently, with the availability of data sets of CT images, methods based on big-data priors have gained interest, such as dictionary learning-based techniques [8]. The dictionary can be either pre-learned from training data, or adaptively learned with the reconstruction. In particular, the synthesis dictionary learning approaches represent a signal or image patch as a sparse linear combination of the atoms or columns of a learned dictionary, and have obtained promising results in many applications [9–11]. However, the dictionary learning based MBIR approaches are often computationally expensive due to expensive sparse coding (where typically NP-hard problems are optimized for estimating sparse coefficients). Different from synthesis dictionary learning, sparsifying transform (a generalized analysis dictionary model) learning techniques efficiently adapt an

operator to approximately sparsify signals in transform domains, and the corresponding transform sparse coding problem can be solved exactly and cheaply by thresholding [12]. Sparsifying transform learning techniques including using doubly-sparse transforms and unions of transforms have been applied to image reconstruction and obtained promising results [13–15].

Very recently, there has been growing interest in deep learning approaches for medical imaging problems [16–21]. In the LDCT image reconstruction field, typical deep learning methods learn the reconstruction mapping from large datasets of pairs of (low-dose and regular-dose) scans. These methods include image-domain learning, sensor-domain learning, and hybrid-domain learning. For example, a particular image-domain learning approach is the FBP-ConvNet scheme [19] that solves the normal-convolutional inverse problems by applying a (learned) CNN after the direct inversion that encapsulates the system physics. The image-domain learning approaches can have many variations. For example, instead of directly working in the image domain, one can transform the images to a specific domain and learn in such domain the relationship between training pairs. Kang et al. [20] designed a neural network that learns a mapping between contourlet transform coefficients of the low-dose input and its high-dose counterpart. This work was later extended to learn a wavelet domain residual network (WavResNet) [21].

In the sensor-domain deep learning category, Würfl et al. [22] proposed an end-to-end neural network for low-dose CT that maps the sinogram to the reconstructed image by mapping the filtered back-projection algorithm to a basic neural network. This allows one to take into account the artifacts in the sensor-domain, e.g., the scatter and beam-hardening artifacts, and compensate them in the learning process. Another framework named the Automated Transform by Manifold Approximation (AUTOMAP) [23] learns a direct mapping from the measurement domain to image domain. However, due to the high memory requirements for storing fully connected layers, it is challenging for AUTOMAP to handle large scale reconstruction tasks such as CT image reconstruction. Hybrid-domain learning approaches exploit data-fidelity terms in the neural network architecture. The Learned ISTA (LISTA) [24] was one of the earliest work of this kind. LISTA unfolds the iterative soft-thresholding (ISTA) algorithm [25], and learns the weight matrices and the sparsifying soft-thresholding operator. Later, Yang et al. [26] proposed an ADMM-Net which unfolds the alternating direction method of multipliers algorithm for image reconstruction. Each step of the algorithm is mapped to a neural network module. This idea was then extended to a learnable primal-dual approach based CNN [27]. These methods fall in the class of physics-driven deep learning methods [28–30]. Hybrid-domain approaches also include a type that applies a plug-and-play model. He et al. [31] applied the plug-and-play model to the ADMM algorithm and unfolded it into a deep reconstruction network, so that each network module is learnable and replaceable.

Most deep learning algorithms are learned in a supervised manner (using task-specific cost functions) and require large training sets. However, in CT imaging, it is often difficult to acquire large datasets of training image pairs. Even though in the AAPM X-ray CT Low-Dose Grand Challenge, both regular-dose and the matched quarter-dose images were provided, only the regular-dose images were reconstructed from real scans, while the matched quarter-dose images were synthesized by adding noise to the regular-dose sinogram data. Therefore, training with smaller number of paired data (and yet generalizing) or without reference data is highly conducive for CT image reconstruction. Moreover, different machine learning (as well as conventional) approaches such as dictionary or transform learning and deep learning use different types of big-data priors and are advantageous in different ways. For example, transform learning approaches learn general image properties and features in an unsupervised or model-based manner, and can easily and effectively adapt to specific image instances.

In this work, we propose a new image reconstruction framework for LDCT dubbed Supervised-UnsuPERvised (SUPER) learning. The algorithm architecture involves interconnected supervised (deep network) and unsupervised (iterative reconstruction) modules over many layers. The architecture enables effectively leveraging different kinds of big data learned priors for CT reconstruction. For example, we used FBPConvNet [19] as the supervised (deep) learned module and PWLS-ULTRA [14] as the unsupervised module with a pre-learned union of transforms, which provided both high quality image reconstruction and image-adaptive clustering. The proposed SUPER learning used relatively small training sets and dramatically outperformed both deep learning and transform/dictionary learning by effectively combining task-specific and image or instance-specific adaptivity. The proposed framework is generalizable to include various constituent modules including non-learning based algorithms, as shown in our experiments.

2. Proposed Model and Algorithms

Here, we present the proposed reconstruction model, its motivations and interpretations, example architectures, and training method.

2.1. Overview of the SUPER Model

We propose a novel efficient physics-driven learning framework for CT reconstruction that effectively combines the benefits of supervised (deep) learning and unsupervised
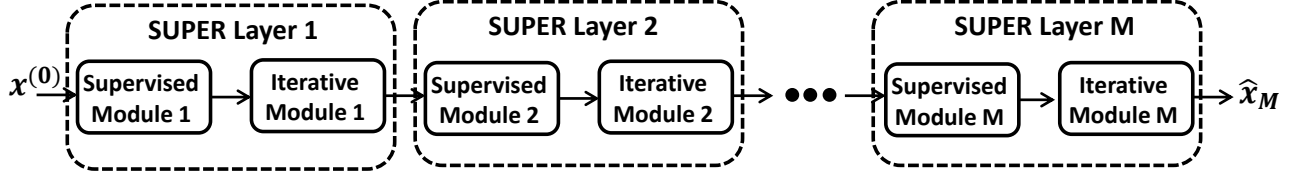
Figure 1: Overall structure of the proposed reconstruction framework.

iterative reconstruction methods. The proposed reconstruction architecture takes an initial image as input and processes it through multiple "super" layers (Fig. 1). Each super layer consists of a network learned in a supervised manner (*supervised module*) and an iterative reconstruction method (*iterative module*) in sequence. The supervised module is different in each super layer, i.e., the weights in the supervised module are not shared among the super layers. Importantly, this module is learned in a supervised manner (e.g., to minimize reconstruction error) to remove artifacts and noise. The iterative module on the other hand iteratively optimizes a regularized image reconstruction problem using image-adaptive priors or regularizers (e.g., the patches of the underlying image can be clustered and sparsified in a learned union of transforms or dictionaries [14]). The iterative algorithm is run for a fixed number of iterations in each super layer.

While the supervised module removes image noise and artifacts using a single learned network, the iterative module could adapt various image-specific features in an MBIR setup to further improve image quality and remove artifacts. Importantly, the iterative module is not learned in a supervised manner. The SUPER model in Fig. 1 is flexible and could use various architectures for the supervised module (e.g., FBPConvNet [19], WavResNet [21], etc.) and a variety of iterative data-adaptive methods (e.g., PWLS-ULTRA [14]). The model could be potentially used in a variety of imaging as well as other applications.

2.2. Interpretations and Generalization

The SUPER model enables combining different kinds of machine learned models and priors in a common reconstruction framework. While the supervised module could be a deep convolutional network learned from a big dataset to optimize task-specific performance metrics, the iterative module could exploit models learned from images using criteria such as sparsity, manifold properties, etc. For example, an operator could be pre-learned from CT images or patches to approximately sparsify them (a.k.a. transform learning [12]) and used to construct the regularizer for the iterative module. Such image-based learned operators are not typically task-sensitive and often generalize readily to different settings and can help delineate or reconstruct various image features.

The proposed SUPER model can also help combine global adaptivity and image-specific adaptivity to obtain the best of both worlds. While each supervised module is learned from a dataset and fixed during reconstruction, the iterative module could optimize novel and specific features for each image being reconstructed (during training and testing) and thus capture the diversity of images and enable highly adaptive reconstructions. For example, during training and testing, the iterative module could cluster image patches differently for each image [14], or even learn novel models such as dictionaries for each image [32].

Another interpretation of the SUPER model arises from the perspective of iterative reconstruction. Many recent state-of-the-art MBIR schemes involve complex nonconvex optimization and priors, wherein the initialization of the algorithm is typically quite important, and better initializations can lead to better reconstructions. In the SUPER model, the iterative module is "initialized" with a different image (i.e., output of the corresponding supervised module) in each super layer. If the output of the supervised module improves in quality over layers, the iterative module will see increasingly better initializations and could thus provide better quality outputs over layers. Moreover, the parameters of the iterative module could also be varied over layers to provide optimal bias-noise trade-offs. Thus, the SUPER model could be viewed as minimizing nonconvex costs in sequence with better initializations and parameters.

The proposed SUPER model for LDCT reconstruction can be generalized to incorporate a variety of iterative and MBIR techniques in the iterative modules. For example, conventional techniques such as PWLS-EP (edge-preserving hyperbola regularizer) [33] could be used in the iterative module. PWLS-EP is a non-adaptive method that penalizes the differences between neighboring pixels in the reconstruction. We show later that combining PWLS-EP with supervised learning in the SUPER model boosts the performance of both methods. Note that we do not run the PWLS-EP modules to near convergence (as it involves a strictly convex problem and a unique minimum) but only for multiple iterations.

2.3. Examples of SUPER Architectures

We now discuss some example SUPER models and their properties. To illustrate the proposed approach, in this work, we focused on the recent FBPConvNet [19] for the supervised module. For the iterative module, we chose the conventional PWLS-EP approach that uses a hand-crafted prior (edge-preserving hyperbola regularizer) as well as the

learning and clustering-based PWLS-ULTRA [14]. Our experiments later show that combining such supervised and iterative methods improves image quality over the constituent methods. In the following, we further discuss the chosen architectures.

2.3.1 Supervised Module

We work with FBPConvNet, which is a CNN-based image-domain denoising architecture, originally designed for sparse-view CT. The backbone of FBPConvNet is a U-net [34] like CNN that takes noisy images reconstructed by the FBP method (from low-dose scans) as input. The neural network is trained so that its outputs closely match the reference high-quality (true) images, e.g., in an ℓ_2 norm or root mean squared error (RMSE) sense.

The traditional U-net uses a multilevel decomposition, and employs a dyadic scale decomposition based on max pooling. Thus, the effective filter size in the middle layers is larger than that in the early and late layers. This scheme is important because the filters corresponding to the Hessian of the data fidelity term in (1) may have noncompact support. Similar to U-net, FBPConvNet employs multichannel filters, which is the standard approach in CNNs, to increase the capacity of the network. Compared with the traditional U-net, FBPConvNet adopts a residual learning strategy to learn the difference between the input and output.

2.3.2 Iterative Module

The iterative module optimizes an MBIR problem that estimates the linear attenuation coefficients $\mathbf{x} \in \mathbb{R}^{N_p}$ from the measurements $\mathbf{y} \in \mathbb{R}^{N_d}$. The typical PWLS approach involves a cost function of the following form:

$$\mathbf{x} = \arg\min_{\mathbf{x} \geq \mathbf{0}} \|\mathbf{y} - \mathbf{A}\mathbf{x}\|_{\mathbf{W}}^2 + \beta \mathsf{R}(\mathbf{x}), \qquad (1)$$

where $\mathbf{A} \in \mathbb{R}^{N_d \times N_p}$ is the CT system matrix, $\mathbf{W}$ is the weighting matrix related to the measurements (capturing measurement statistics), $\mathsf{R}(\mathbf{x})$ is the regularizer, and β is a positive scalar controlling the balance between the data-fidelity term and the regularizer. In this paper, we used the unsupervised learning-based PWLS-ULTRA as well as the conventional PWLS-EP for (1).

For PWLS-EP, the regularizer $\mathsf{R}(\mathbf{x})$ can be written as $\mathsf{R}(\mathbf{x}) = \sum_{j=1}^{N_p} \sum_{k \in N_j} \kappa_j \kappa_k \varphi(x_j - x_k)$, where x_j is the jth pixel of $\mathbf{x}$, N_j is the neighborhood of the jth pixel, and κ_j and κ_k are analytically determined weights that encourage resolution uniformity [33]. The potential function $\varphi(t)$ is often chosen as $\varphi(t) = \delta^2(|t/\delta| - \log(1 + |t/\delta|))$, for $\delta > 0$. PWLS-EP enforces approximate sparsity of gradients of the image, a non-adaptive prior.

PWLS-ULTRA pre-learns a union of sparsifying transforms from a dataset of image patches, and uses the learned model during reconstruction [14]. The formulation for

learning the union of sparsifying transforms is as follows:

$$\min_{\{\mathbf{\Omega}_k, \mathbf{Z}_i, C_k\}} \sum_{k=1}^{K} \sum_{i \in C_k} \{\|\mathbf{\Omega}_k \mathbf{X}_i - \mathbf{Z}_i\|_2^2 + \eta \|\mathbf{Z}_i\|_0\}$$
$$+ \sum_{k=1}^{K} \lambda_k \mathbf{Q}(\mathbf{\Omega}_k), \text{ s.t. } C_k \in \mathcal{G}. \qquad (2)$$

Here, $\mathbf{\Omega}_k$, C_k, and $\mathbf{Z}_i$ represent the learned transform for the kth class, the set of indices of patches belonging to the kth class, and the sparse coefficient of the ith training signal or patch $\mathbf{X}_i$ (N training signals assumed), respectively. Each signal is grouped with a corresponding best matched sparsifying transform in (2). The set $\mathcal{G}$ is the set of all possible partitions of $[1 : N]$ into K disjoint subsets. To avoid trivial solutions for $\mathbf{\Omega}_k$, the penalty terms $\mathbf{Q}(\mathbf{\Omega}_k) \triangleq \|\mathbf{\Omega}_k\|_F^2 - \log|\det \mathbf{\Omega}_k|$ for $1 \leq k \leq K$ are used that also control the condition number of the transforms. Parameters η and $\lambda_k = \lambda_0 \sum_{i \in C_k} \|\mathbf{X}_i\|_2^2$ are positive weighting factors with $\lambda_0 > 0$ [14]. Problem (2) is solved efficiently using alternating optimization [14].

With the pre-learned transforms $\{\mathbf{\Omega}_k\}$, the regularizer $\mathsf{R}(\mathbf{x})$ for image reconstruction is as follows:

$$\mathsf{R}(\mathbf{x}) \triangleq \min_{\{\mathbf{z}_j, C_k\}} \sum_{k=1}^{K} \sum_{j \in C_k} \tau_j \left\{ \|\mathbf{\Omega}_k \mathbf{P}_j \mathbf{x} - \mathbf{z}_j\|_2^2 + \gamma^2 \|\mathbf{z}_j\|_0 \right\},$$

where $\{\tau_j\}$ are patch-based weights to encourage uniform spatial resolution or uniform noise (see [14]), $\mathbf{P}_j \in \mathbb{R}^{l \times N_p}$ is a patch extraction operator that extracts the jth patch from $\mathbf{x}$. $\mathbf{z}_j$ is the sparse coefficient for the jth patch, and $\gamma > 0$ is a parameter controlling sparsity.

The PWLS-ULTRA problem is efficiently solved by alternating between updating $\mathbf{x}$ (*image update step*), and solving for $\{\mathbf{z}_j, C_k\}$ (*sparse coding and clustering step*). In the image update step, where $\{\mathbf{z}_j, C_k\}$ are fixed, the subproblem is quadratic with non-negativity constraints, and can be solved using fast iterative algorithms such as the relaxed linearized augmented Lagrangian method with ordered-subsets (relaxed OS-LALM) [35, 36]. The sparse coding and clustering step with fixed $\mathbf{x}$ is solved exactly [32], with the optimal class assignment $\hat{k}_j$ for each patch given as

$$\arg\min_{1 \leq k \leq K} \|\mathbf{\Omega}_k \mathbf{P}_j \mathbf{x} - H_\gamma(\mathbf{\Omega}_k \mathbf{P}_j \mathbf{x})\|_2^2 + \gamma^2 \|H_\gamma(\mathbf{\Omega}_k \mathbf{P}_j \mathbf{x})\|_0.$$

The corresponding optimal $\hat{\mathbf{z}}_j = H_\gamma(\mathbf{\Omega}_{\hat{k}_j} \mathbf{P}_j \mathbf{x})$, where $H_\gamma(\cdot)$ is the hard-thresholding operator that sets elements smaller than γ to zero. The hard-thresholding can be viewed as the non-smooth nonlinearity in PWLS-ULTRA. The clustering could vary from image to image and iteration to iteration in the PWLS-ULTRA algorithm.

2.4. Training and Implementation

We propose to train the SUPER model layer-by-layer (sequentially) from a dataset of pairs of low-dose and

regular-dose CT measurements. For example, the FBP method can be used to obtain reconstructions from the measurements. The initial low-dose reconstructed images are then used as inputs to the first supervised module, which is trained to minimize the reconstruction error (RMSE) at its output, with respect to the regular-dose reconstructions. The initial images are then passed through the trained network, following which the iterative algorithm in the first iterative module is run for each training image (in parallel) to produce iterative reconstructions. The iterative reconstructions serve as inputs to the subsequent supervised model, which is trained to minimize reconstruction error. The subsequent supervised modules are thus learned sequentially.

Once trained, the SUPER model is readily implemented for test data by passing initial reconstructions sequentially through the supervised learned networks and iterative algorithms.

3. Experiments

Here, we first describe our experimental setup, training procedures, and evaluation metrics. Then we present results for the learned SUPER-ULTRA model, and compare these with those obtained by each individual module of SUPER, i.e., FBPConvNet and PWLS-ULTRA. We also tested the generalized SUPER model that replaces the unsupervised learning-based PWLS-ULTRA with the non-adaptive PWLS-EP. This scheme is dubbed FBPConvNet+EP.

3.1. Experimental Setup

We used regular-dose CT images of two patients from the Mayo Clinics dataset established for "*the 2016 NIH-AAPM-Mayo Clinic Low Dose CT Grand Challenge*" [37], to evaluate the performance of the proposed SUPER learning. The two patient datasets (L067 and L096) contain 224 and 330 real in-vivo slice images, respectively. We simulated low-dose CT measurements $\mathbf{y}_l$ from the provided regular-dose images with GE 2D LightSpeed fan-beam CT geometry corresponding to a monoenergetic source. We projected the regular-dose images $\mathbf{x}^*$ to sinograms and added Poisson and additive Gaussian noise to them as follows:

$$\mathbf{y}_{l_i} = \text{Poisson}\{I_0 e^{-[\mathbf{A}\mathbf{x}^*]_i}\} + \mathcal{N}\{0, \sigma^2\}. \qquad (3)$$

We chose $I_0 = 1 \times 10^5$ photons per ray and $\sigma = 5$ in our experiments. We approximated elements of the diagonal weighting matrix $\mathbf{W}$ of the data-fidelity in (1) by $\frac{\mathbf{y}_{l_i}^2}{\mathbf{y}_{l_i}+\sigma^2}$ [38]. The images are of size 512×512 at a resolution of $0.9766 \text{ mm} \times 0.9766 \text{ mm}$, when reconstructed using the FBP method. These reconstructed low-dose FBP images were paired with their corresponding regular-dose CT images for training the SUPER model.

3.2. Training Procedures

In our experiments, the total number of training image pairs was 100, of which 50 were chosen from patient L067 and 50 from patient L096. The image set was used to train the networks of FBPConvNet, FBPConvNet+EP, and SUPER-ULTRA. Fig. 2 shows some example (regular-dose or reference) images from the training and testing datasets. Different body parts are included in the datasets. In our experiments, we used the default network architecture in the FBPConvNet public implementation. For PWLS-ULTRA, we pre-learned a union of five sparsifying transforms (corresponding to five classes) from twelve (regular-dose) slices that include three slices each from four patients (L109, L143, L192, L506).
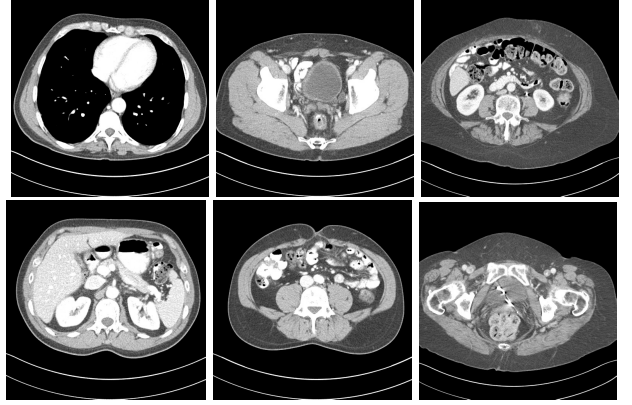

Figure 2: Example CT images in the training (top row) and testing (bottom row) datasets. The display window is [800 1200] HU.

We used a GTX Titan GPU graphic processor for training and testing. The union of transforms was learned efficiently using similar parameters as in [14]. For SUPER-ULTRA and FBPConvNet+EP, the training time was about 46 hours and 13 hours, respectively, for 15 super layers. Since the iterative reconstruction modules are relatively computationally expensive, we chose 15 super layers to balance reconstruction quality and computational time. The training hyper-parameters for the CNN part of FBPConvNet were set as follows for the various models: the learning rate decreased logarithmically from 0.001 to 0.0001; the batch-size was 1; and the momentum was 0.99. We initialized the filters in the various networks during training with i.i.d. random Gaussian entries with zero mean and variance 0.005, and initialized the bias with zero vectors.

For training FBPConvNet+EP, we ran 10 epochs of FBPConvNet training (using ADAM) in each super layer to ensure that the learned networks capture layer-wise features. For the constituent PWLS-EP modules, we ran 4 iterations of the relaxed OS-LALM algorithm with 4 subsets, and set $\delta = 20$ HU and the regularization parameter $\beta = 2^{15}$. For training SUPER-ULTRA, we ran 10 epochs of FBPCon-
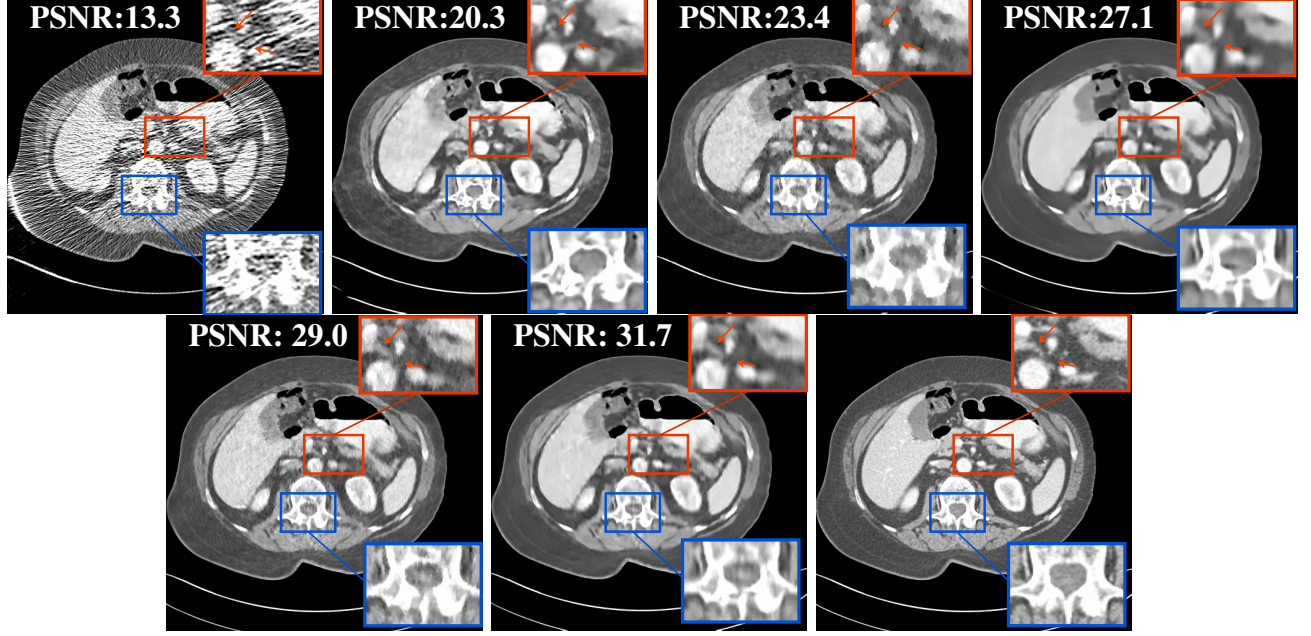
Figure 3: Reconstructed testing image (Test #4, from patient L096) obtained by FBP, FBPConvNet, PWLS-EP, PWLS-ULTRA, FBPConvNet + EP, SUPER-ULTRA, and regular-dose FBP. The display window is [800 1200] HU.
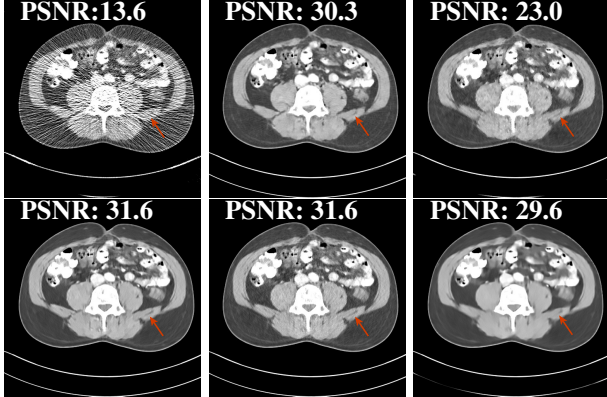

Figure 4: Comparison of Test #2 for FBP, FBPConvNet, PWLS-EP, PWLS-ULTRA, FBPConvNet + EP, and SUPER-ULTRA (clockwise top left). The display window is [800 1200] HU.

vNet training in each super layer along with 4 outer iterations of PWLS-ULTRA with 5 inner iterations of the image update step that used the relaxed OS-LALM algorithm with 4 subsets. We set $\beta = 5 \times 10^3$ and $\gamma = 20$ for the PWLS-ULTRA module.

We compared our learned models with the iterative schemes PWLS-EP and PWLS-ULTRA. Since FBPConvNet+EP and SUPER-ULTRA already include the learned FBPConvNet modules, we used smaller regularization parameters for them compared to the usual or stand-alone PWLS-EP and PWLS-ULTRA, which worked well in our experiments. For the stand-alone PWLS-EP iterative recon-

struction algorithm that solves a convex problem, β and δ were set as 2^{16} and 20, respectively, and we ran 100 iterations of the OS-LALM algorithm to ensure convergence. For the stand-alone PWLS-ULTRA, wherein the optimization problem is nonconvex, we set β and γ as 10^4 and 25, respectively, and ran 1000 iterations of the alternating algorithm to ensure convergence. The above parameters provided optimal image quality in our experiments. Both PWLS-EP and PWLS-ULTRA were initialized with the simple FBP reconstructions.

3.3. Evaluation Metrics

To quantitatively evaluate the performances of the various reconstruction models, we chose three classic metrics, namely RMSE, peak signal to noise ratio (PSNR), and structural similarity index measure (SSIM) [39]. RMSE in Hounsfield units (HU) is defined as RMSE=$\sqrt{\sum_{i=1}^{N_p} \left(\hat{x}_i - x_i^* \right)^2 / N_p}$, where $\mathbf{x}^*$ is the reference regular-dose CT image, $\hat{\mathbf{x}}$ is the reconstruction, and N_p is the number of pixels. We computed PSNR in decibels (dB).

3.4. Results and Comparisons
3.4.1 Visual Results

We applied the various methods to six testing slices (three slices from L067 and three slices from L096) indexed as Test #1 to Test #6. Figs. 3 and 4 show the reconstructions of Test #4 (from patient L096) and Test #2 (from patient L067), respectively, using different methods. Clearly, the traditional FBP reconstruction shows severe artifacts, while
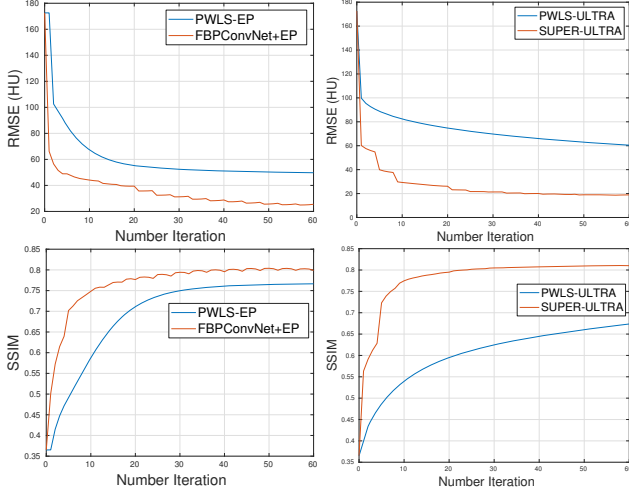
Figure 5: RMSE (first row) and SSIM (second row) values for Test #4 over the iterations of PWLS-EP, FBPConvNet+EP, PWLS-ULTRA, and SUPER-ULTRA.

the results obtained by the other methods have much better image quality. The stand-alone optimal PWLS-EP reconstruction still contains noise such as in the central soft-tissue area (Fig. 3). The (unsupervised) learning-based PWLS-ULTRA removes noise, but the edges of soft-tissues are more blurry in the results. The supervised learning-based FBPConvNet achieves better trade-off between resolution and noise compared to PWLS-EP and PWLS-ULTRA, but many small structures were missed or distorted. With the same training data as FBPConvNet, the learned FBPConvNet+EP and SUPER-ULTRA models provide much better reconstructions. However, FBPConvNet+EP suffers from some streak artifacts in the central area (Fig. 3) as well as noise generally, and SUPER-ULTRA mitigates these artifacts. In both Figs. 3 and 4, SUPER-ULTRA that combines supervised learned networks and transform learning-based iterative reconstructions achieved the best overall visual quality. [1]

3.4.2 Quantitative Results

Fig. 5 shows the RMSE and SSIM evolution for the stand-alone PWLS-EP and PWLS-ULTRA along with those for FBPConvNet+EP and SUPER-ULTRA. The latter two involve 15 super layers with 4 (outer) iterations in the iterative module per layer, and thus, the RMSE and SSIM evolution is plotted over these individual iterations. For PWLS-EP and PWLS-ULTRA, we show the evolution of the metrics over 60 iterations. FBPConvNet+EP and SUPER-ULTRA clearly achieve much lower RMSE values and higher SSIM values over layers than PWLS-EP and PWLS-ULTRA. The plots also show faster convergence for the SUPER models.

Table. 1 shows the RMSE, PSNR, and SSIM values

---

[1] Additional comparisons between reconstructions for the other testing slices are included in the supplement.
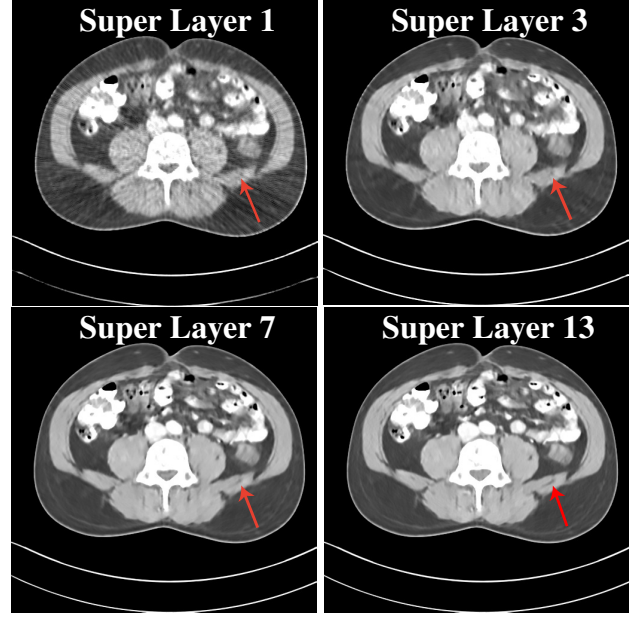

Figure 6: Outputs of 1st, 3rd, 7th, and 13th SUPER-ULTRA layers with display window [800 1200] HU. The reconstruction is visually converging.

with various methods for the testing slices. The proposed SUPER-ULTRA typically achieves significant improvements in RMSE, SSIM, and PSNR over the other methods for all testing slices. Importantly, FBPConvNet+EP and SUPER-ULTRA perform much better than PWLS-EP and PWLS-ULTRA, respectively, and also provide more promising results than FBPConvNet, demonstrating that the combination of the supervised module and iterative modules in the SUPER model works well and outperforms the individual modules.

3.4.3 Visual Quality over SUPER layers

Fig. 6 shows the output of SUPER-ULTRA after different numbers of SUPER layers. The initial FBP image and the final output (after 15 layers) were shown in Fig. 4. The first several layers of SUPER-ULTRA mainly remove severe noise and artifacts, while the later layers recover some structural details.

3.4.4 Behavior of the ULTRA model in the SUPER architecture

Next, to better illustrate the image-adaptive learned clustering in the SUPER-ULTRA model, Fig. 7 shows pixel-level clustering results from the last super layer for test slice #1. Since the ULTRA modules cluster image patches into specific classes, we cluster each pixel here using a majority vote among the patches overlapping the pixel. Class 1 contains most of the more uniform soft-tissues; Classes 2, 3, and 5 contain many oriented edges (e.g., at 45-degree and 135-degree orientation); and class 4 contains most of the

Table 1: PSNR, RMSE, and SSIM of reconstructed test images for different methods.

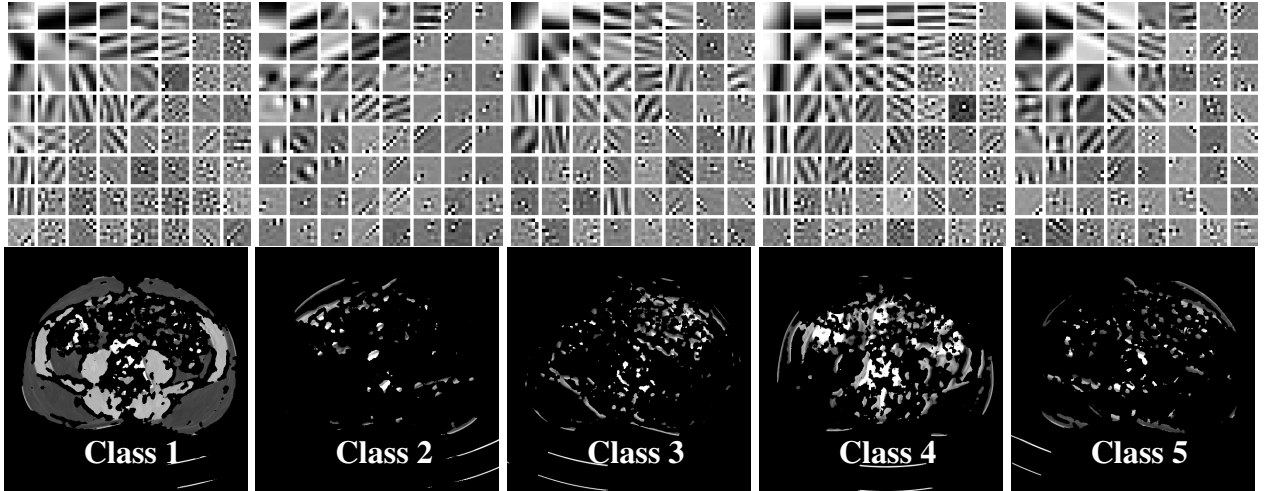|			FBP	FBPConvNet	PWLS-EP	PWLS-ULTRA	FBPConvNet+EP	SUPER-ULTRA
L067	Test #1	PSNR	11.0	29.8	23.5	29.3	30.5	**30.9**
		RMSE	245.5	26.2	54.1	28.3	24.5	**23.3**
		SSIM	0.29	0.74	0.73	0.74	0.77	**0.77**
	Test #2	PSNR	13.6	30.3	23.0	29.6	31.6	**31.6**
		RMSE	170.3	22.4	52.2	24.5	19.5	**19.3**
		SSIM	0.40	0.79	0.78	0.79	0.81	**0.81**
	Test #3	PSNR	9.5	19.9	24.4	28.6	**32.2**	32.1
		RMSE	299.5	81.7	47.9	29.6	**19.6**	19.8
		SSIM	0.24	0.56	0.69	0.69	0.72	**0.72**
L096	Test #4	PSNR	13.3	20.3	23.4	27.1	29.0	**31.7**
		RMSE	172.6	70.7	48.7	31.9	25.5	**18.9**
		SSIM	0.37	0.67	0.77	0.79	0.80	**0.81**
	Test #5	PSNR	9.3	29.7	23.3	25.7	30.6	**30.8**
		RMSE	304.4	25.8	53.8	40.6	23.3	**22.7**
		SSIM	0.20	0.71	0.70	0.70	0.74	**0.75**
	Test #6	PSNR	9.7	27.6	23.6	28.1	30.7	**32.3**
		RMSE	274.2	32.3	48.6	29.0	21.5	**17.8**
		SSIM	0.23	0.66	0.73	0.74	0.75	**0.76**

Figure 7: Pixel-level clustering results in Test #2. The top row shows the transforms, with the transform rows shown as 8×8 patches. The bottom row shows the clustering results of SUPER-ULTRA with display window [800 1200] HU.

vertical edges and some horizontal edges as well as most of the bones. The latter classes help provide sharper SUPER-ULTRA reconstructions. The pre-learned transforms corresponding to each class are also shown in Fig. 7, and contain various directional and edge-like features.

4. Conclusions

This paper presented a new framework that combined supervised learned networks and unsupervised iterative algorithms for low-dose CT reconstruction. The proposed SUPER framework effectively combines various kinds of priors and learning methods. In particular, we studied SUPER-ULTRA that combines (supervised) deep learning (FBP-ConvNet) and the recent iterative (unsupervised) PWLS-ULTRA, as well as FBPConvNet+EP (or SUPER-EP). Both methods showed better performance and faster convergence compared to their individual modules. FBPConvNet+EP substantially improved the performance of PWLS-EP, while SUPER-ULTRA typically performed the best by effectively leveraging deep learning and transform learning. While SUPER model learning can exploit a variety of architectures and algorithms for the supervised and iterative modules, a more detailed study of various such architectures is left for future work. We also plan to explore layer-dependent parameter selection for the iterative modules to further improve performance in future work.

References

[1] L. C. Feldkamp, L. A .and Davis and J. W. Kress. Practical cone-beam algorithm. *J. Opt. Soc. Amer. A, Opt. Image Sci.*, 1(6):612–619, 1984.

[2] K. Imai, M. Ikeda, Y. Enchi, and T. Niimi. Statistical characteristics of streak artifacts on CT images: Relationship between streak artifacts and ma s values. *Med. Phys.*, 36(2):492–499, 2009.

[3] H. Zhang, J. Ma, J. Wang, Y. Liu, H. Lu, and Z. Liang. Statistical image reconstruction for low-dose CT using nonlocal means-based regularization. *Computerized Medical Imaging and Graphics*, 38(6):423–435, 2014.

[4] J. Nuyts, B. De Man, J. A. Fessler, W. Zbijewski, and F. J. Beekman. Modelling the physics in the iterative reconstruction for transmission computed tomography. *Phys. Med. Biol.*, 58(12):R63, 2013.

[5] Jeffrey A Fessler. Penalized weighted least-squares image reconstruction for positron emission tomography. *IEEE transactions on medical imaging*, 13(2):290–300, 1994.

[6] 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(11):4526–44, November 2007.

[7] M. Beister, D. Kolditz, and W. A Kalender. Iterative reconstruction methods in X-ray CT. *Physica Medica: European Journal of Medical Physics*, 28(2):94–108, 2012.

[8] Q. Xu, H. Yu, X. Mou, L. Zhang, J. Hsieh, and G. Wang. Low-dose X-ray CT reconstruction via dictionary learning. *IEEE Trans. Med. Imag.*, 31(9):1682–97, September 2012.

[9] S. Ravishankar and Y. Bresler. MR image reconstruction from highly undersampled k-space data by dictionary learning. *IEEE Trans. Med. Imag.*, 30(5):1028–1041, 2010.

[10] J. Mairal, M. Elad, and G. Sapiro. Sparse representation for color image restoration. *IEEE Trans. Im. Proc.*, 17(1):53–69, 2007.

[11] M. Elad and M. Aharon. Image denoising via sparse and redundant representations over learned dictionaries. *IEEE Trans. Im. Proc.*, 15(12):3736–3745, 2006.

[12] S. Ravishankar and Y. Bresler. Learning sparsifying transforms. *IEEE Trans. Signal Process.*, 61(5):1072–1086, 2013.

[13] S. Ravishankar and Y. Bresler. Learning doubly sparse transforms for images. *IEEE Trans. Im. Proc.*, 22(12):4598–4612, 2013.

[14] 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(6):1498–1510, 2018.

[15] S. Ye, Y. Ravishankar, S.and Long, and J. A. Fessler. SPUL-TRA: Low-dose CT image reconstruction with joint statistical and learned image models. *IEEE Trans. Med. Imag.*, 2019. DOI: 10.1109/TMI.2019.2934933.

[16] G. Yang, S. Yu, H. Dong, G. Slabaugh, P. L. Dragotti, X. Ye, F. Liu, S. Arridge, J. Keegan, Y. Guo, and D. Firmin. DA-GAN: deep de-aliasing generative adversarial networks for fast compressed sensing MRI reconstruction. *IEEE Trans. Med. Imag.*, 37(6):1310–1321, 2017.

[17] S. Yu, H. Dong, G. Yang, G. Slabaugh, P. L. Dragotti, X. Ye, F. Liu, S. Arridge, J. Keegan, D. Firmin, and Y. Guo. Deep de-aliasing for fast compressive sensing MRI. *arXiv preprint arXiv:1705.07137*, 2017.

[18] J. Schlemper, G. Yang, P. Ferreira, A. Scott, L. McGill, Z. Khalique, M. Gorodezky, M. Roehl, J. Keegan, D. Pennell, D. Firmin, and D. Rueckert. Stochastic deep compressive sensing for the reconstruction of diffusion tensor cardiac MRI. In *Med. Image Comput. Comput.-Assist. Interv.*, pages 295–303. Springer, 2018.

[19] K. H. Jin, M. T. McCann, E. Froustey, and M. Unser. Deep convolutional neural network for inverse problems in imaging. *IEEE Trans. Im. Proc.*, 26(9):4509–22, 2017.

[20] E. Kang, J. Min, and J. C. Ye. A deep convolutional neural network using directional wavelets for low-dose X-ray CT reconstruction. *Med. Phys.*, 44(10):e360–e375, 2017.

[21] E. Kang, W. Chang, J. Yoo, and J. C. Ye. Deep convolutional framelet denoising for low-dose CT via wavelet residual network. *IEEE Trans. Med. Imag.*, 37(6):1358–1369, 2018.

[22] T. Würfl, F. C. Ghesu, V. Christlein, and A. Maier. Deep learning computed tomography. In *Med. Image Comput. Comput. Assist. Interv.*, pages 432–440. Springer, 2016.

[23] B. Zhu, J. Z. Liu, S. F. Cauley, B. R. Rosen, and M. S. Rosen. Image reconstruction by domain-transform manifold learning. *Nature*, 555(7697):487, 2018.

[24] K. Gregor and Y. LeCun. Learning fast approximations of sparse coding. In *Proceedings of the 27th International Conference on International Conference on Machine Learning*, pages 399–406. Omnipress, 2010.

[25] A. Beck and M. Teboulle. A fast iterative shrinkage-thresholding algorithm for linear inverse problems. *SIAM J. Imaging Sci.*, 2(1):183–202, 2009.

[26] Y. Yang, J. Sun, H. Li, and Z. Xu. Deep ADMM-Net for compressive sensing MRI. In *Proceedings of the 30th International Conference on Neural Information Processing Systems*, pages 10–18. Curran Associates Inc., 2016.

[27] J. Adler and O. Öktem. Learned primal-dual reconstruction. *IEEE Trans. Med. Imag.*, 37(6):1322–1332, 2018.

[28] S. Ravishankar, I. Y. Chun, and J. A. Fessler. Physics-driven deep training of dictionary-based algorithms for MR image reconstruction. In *2017 51st Asilomar Conference on Signals, Systems, and Computers*, pages 1859–1863, 2017.

[29] S. Ravishankar, A. Lahiri, C. Blocker, and J. A. Fessler. Deep dictionary-transform learning for image reconstruction. In *2018 IEEE 15th International Symposium on Biomedical Imaging (ISBI 2018)*, pages 1208–1212, 2018.

[30] Y. Chun and J. A. Fessler. Deep BCD-Net using identical encoding-decoding CNN structures for iterative image

recovery. In *2018 IEEE 13th Image, Video, and Multidimensional Signal Processing Workshop (IVMSP)*, pages 1–5, 2018.

[31] J. He, Y. Yang, Y. Wang, D. Zeng, Z. Bian, H. Zhang, J. Sun, Z. Xu, and J. Ma. Optimizing a parameterized plug-and-play admm for iterative low-dose CT reconstruction. *IEEE Trans. Med. Imag.*, 38(2):371–382, 2018.

[32] S. Ravishankar and Y. Bresler. Data-driven learning of a union of sparsifying transforms model for blind compressed sensing. *IEEE Trans. Comput. Imag.*, 2(3):294–309, 2016.

[33] 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(2):678–89, February 2015.

[34] O. Ronneberger, P. Fischer, and T. Brox. U-net: convolutional networks for biomedical image segmentation. In *Medical Image Computing and Computer-Assisted Intervention*, pages 234–41, 2015.

[35] H. Nien and J. A. Fessler. Relaxed linearized algorithms for faster X-ray CT image reconstruction. *IEEE Trans. Med. Imag.*, 35(4):1090–8, April 2016.

[36] D. Kim and J. A. Fessler. Generalizing the optimized gradient method for smooth convex minimization. *SIAM J. Optim.*, 28(2):1920–1950, 2018.

[37] C. McCollough. TU-FG-207A-04: Overview of the low dose CT grand challenge. *Med. Phys.*, 43(2):3759–60, 2016.

[38] L. Fu, T. C. Lee, S. M. Kim, A. M. Alessio, P. E. Kinahan, Z. Q. Chang, K. Sauer, M. K. Kalra, and B. De Man. Comparison between pre-log and post-log statistical models in ultra-low-dose CT reconstruction. *IEEE Trans. Med. Imag.*, 36(3):707–720, 2017.

[39] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli. Image quality assessment: from error visibility to structural similarity. *IEEE Trans. Im. Proc.*, 13(4):600–12, April 2004.