DEEP RESIDUAL LEARNING FOR MODEL-BASED ITERATIVE CT RECONSTRUCTION USING PLUG-AND-PLAY FRAMEWORK

Dong Hye Ye[*], *Somesh Srivastava*[†], *Jean-Baptiste Thibault*[†], *Jiang Hsieh*[†], *Ken Sauer*[‡], *Charles Bouman*[*],

[*]School of Electrical and Computer Engineering, Purdue University, West Lafayette, IN, 47907
[†]GE Healthcare Technologies, Waukesha, WI, 53188
[‡]Department of Electrical Engineering, University of Notre Dame, Notre Dame, IN, 46556

ABSTRACT

Model-Based Iterative Reconstruction (MBIR) has shown promising results in clinical studies as they allow significant dose reduction during CT scans while maintaining the diagnostic image quality. MBIR improves the image quality over analytical reconstruction by modeling both the sensor (e.g., forward model) and the image being reconstructed (e.g., prior model). While the forward model is typically based on the physics of the sensor, accurate prior modeling remains a challenging problem. Markov Random Field (MRF) has been widely used as prior models in MBIR due to simple structure, but they cannot completely capture the subtle characteristics of complex images. To tackle this challenge, we generate a prior model by learning the desirable image property from a large dataset. Toward this, we use Plug-and-Play (PnP) framework which decouples the forward model and the prior model in the optimization procedure, replacing the prior model optimization by a image denoising operator. Then, we adopt the state-of-the-art deep residual learning for the image denoising operator which represents the prior model in MBIR. Experimental results on real CT scans demonstrate that our PnP MBIR with deep residual learning prior significantly reduces the noise and artifacts compared to standard MBIR with MRF prior.

Index Terms— Model-Based Iterative CT Reconstruction, Plug-and-Play Framework, Deep Residual Learning

1. INTRODUCTION

Model-Based Iterative Reconstruction (MBIR) has gained increasing attention due to its superiority for low-dose CT reconstruction compared with analytical methods such as filtered-back projection (FBP) [1, 2, 3]. The benefit of MBIR is due to incorporating modeling of physics of CT scanner and desirable image properties into a reconstruction problem. Among many modelings such as the forward projection and electronic noise models, the image prior model plays a significant role in the quality of MBIR. The general approach used in MBIR is to model the image as Markov Random Field (MRF) [4], which penalizes intensity fluctuations in the neighborhood [5, 6, 7]. However, MRF is not sufficient to differentiate the noise-induced fluctuations in the image because there is little prior information available about the underlying object. In this paper, we instead generate the prior model by applying machine learning techniques to a large number of training images.

Even though the data-driven image model can accurately capture the image properties of the underlying object, it is still challenging to integrate it into MBIR as a prior model because there is no explicit formula. To tackle this challenge, we use the Plug-and-Play (PnP) framework [8, 9] which utilizes alternating direction method of multipliers (ADMM) [10, 11] for the MBIR optimization. ADMM splits the state variables in MBIR cost function so as to decouple the forward and prior model terms, resulting in two separate constrained optimization problems. This leads to two completely independent modules for MBIR implementation: one module for reconstruction dependent on the forward model, and the other module for denoising based on the prior model. With this PnP framework, any denoising algorithm using machine learning techniques can be used as a prior model in MBIR.

Then, we use the denoising algorithm based on deep learning as a prior model for MBIR. In such algorithms, the weights of a neural network are trained on large datasets, and the trained neural network is applied to remove noise of the unseen testing image [12, 13, 14, 15, 16]. While only shallow neural networks can be trained in classic approaches due to the high computational time needed, recent advances in network architecture such as deep residual learning allow us to learn much deeper neural networks [17, 18]. These deep residual learning based methods show great success in low-dose CT denoising [19]. Therefore, we adapt deep residual learning based denoising as the prior model for MBIR. Experimental results on real CT scans show that PnP MBIR with deep residual learning prior achieves more accurate reconstruction than standard MBIR with MRF prior because the generated data-driven prior information quickly drives to the fully converged solution for iterative optimization.

The authors acknowledge GE Healthcare for supporting this work.

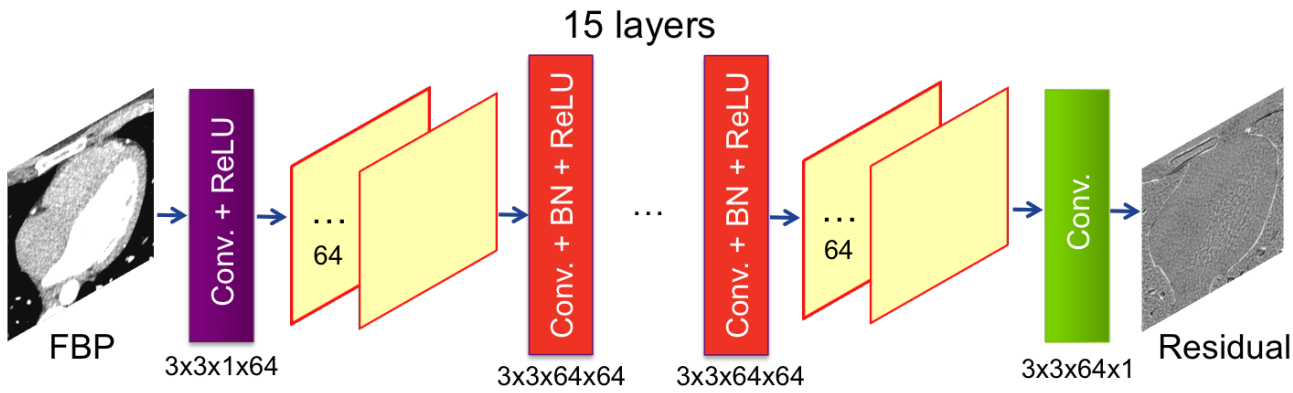


Fig. 1. Network architecture for deep residual learning: Given a noisy FBP image, 17-layer convolutional neural network is constructed to predict a residual image. Note that we use batch normalizaton (BN) and rectified linear unit (ReLU) for efficient training.

2. MBIR WITH DEEP RESIDUAL LEARNING

Let $x \in \mathbb{R}^N$ denote the image and $y \in \mathbb{R}^M$ be the measured CT scan data. In MBIR, the image is reconstructed by solving the *maximum a posteriori* (MAP) estimation problem:

$$\hat{x} = \underset{x}{\operatorname{argmin}} \frac{1}{2}||\mathbf{A}x - y||^2_{\mathbf{W}} + \frac{1}{2\sigma^2}\Phi(x), \tag{1}$$

where $\mathbf{A} \in \mathbb{R}^{M \times N}$ is the system matrix which represents CT scanner geometry, and $\mathbf{W} \in \mathbb{R}^{M \times M}$ is a diagonal weighting matrix which determines the influence of individual measurements according to noise variance. The prior model term $\Phi(x)$ is introduced to stabilize the estimate and enforce desirable image properties in the reconstructed image. The typical prior models such as Markov Random Fields (MRF) penalize local gradients in the image (e.g., $\Phi(x) = \sum_{k \in \mathcal{N}_j} |x_j - x_k|^2$ where $\mathcal{N}_j$ represents local neighborhood of pixel j).

2.1. Plug-and-Play for MBIR

The solution to eq.1 can be achieved using a variety of optimization methods, such as Iterative Coordinate Descent (ICD) [20] and Ordered Subset (OS) [21]. However, these optimization methods requires the gradient computation which constrains the prior model $\Phi(x)$ to be first-order differentiable. Therefore, we instead use the alternating direction method of multipliers (ADMM) [10, 11] which separates the prior model during cost optimization.

Toward this, we first split the variable x into v, resulting in the following MAP estimation:

$$(\hat{x}, \hat{v}) = \underset{\substack{x,v \\ x=v}}{\operatorname{argmin}} \frac{1}{2}||\mathbf{A}x - y||^2_{\mathbf{W}} + \frac{1}{2\sigma^2}\Phi(v). \tag{2}$$

We solve this constrained optimization problem using the augmented Lagrangian method.

$$L(x, v, u) = \frac{1}{2}||\mathbf{A}x - y||^2_{\mathbf{W}} + \frac{1}{2\sigma^2}\Phi(v) + \frac{1}{2\lambda^2}||x - v + u||^2_2, \tag{3}$$

where u is a scaled dual variable and λ is a relative weight parameter.

By first solving x with the fixed v and then solving for v with x fixed in minimizing $L(x, v, u)$, the ADMM iteratively updates each variables as following.

$$\hat{x} = \underset{x}{\operatorname{argmin}} \frac{1}{2}||\mathbf{A}x - y||^2_{\mathbf{W}} + \frac{1}{2\lambda^2}||x - \tilde{x}||^2_2, \tag{4}$$

$$\hat{v} = \underset{v}{\operatorname{argmin}} \frac{1}{2\sigma^2}\Phi(v) + \frac{1}{2\lambda^2}||\tilde{v} - v||^2_2, \tag{5}$$

$$u \leftarrow u + (\hat{x} - \hat{v}), \tag{6}$$

where $\tilde{x} = \hat{v} - u$ and $\tilde{v} = \hat{x} + u$. It is worth noting that eq. 6 is not the optimization over u but the direct update on the dual variable. Even though these procedures are not equivalent to exact minimization of eq. 2, ADMM converges to the approximate solution [9].

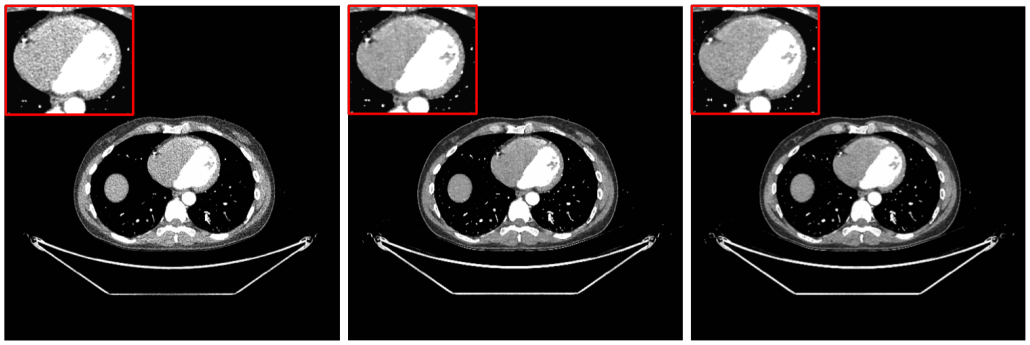
The solution of eq. 4 can be achieved via simple gradient descent algorithm approximating to the closed-form solution.

$$\hat{x} \approx (\mathbf{A}^T\mathbf{W}\mathbf{A} + \frac{1}{\lambda^2}\mathbf{I})^{-1}(\mathbf{A}^T\mathbf{W}y + \frac{1}{\lambda^2}\tilde{x}) \tag{7}$$

For eq. 5, it becomes the classic image denoising problem which can be denoted as the function of $\tilde{v}$.

$$\hat{v} = F(\tilde{v}) = \underset{v}{\operatorname{argmin}} ||\tilde{v} - v||^2_2 + \frac{\lambda^2}{\sigma^2}\Phi(v). \tag{8}$$

Therefore, we can use any image denoising operators which do not have the explicit form of solution to represent the prior model. We call this framework as Plug-and-Play (PnP) because we can incorporate the state-of-the-art image denoising method into MBIR as a prior model.



(a) FBP (b) Deep Residual Learning (c) Ground-truth

Fig. 2. Example of image denoising based on deep residual learning: (a) FBP input, (b) Denoised FBP by deep residual learning, (c) Ground-truth. The display window of the intensity map is $[800, 1200]$ HU. Cardiac ROI is zoomed in the red rectangle. Note that deep residual learning significantly reduces the noise and enhances the resolution compared with FBP while generating very close image to the fully converged ground-truth image.

2.2. Deep Residual Learning for Image Denoising

The goal of denoising is to find the mapping from the noisy input image to the latent clean image (e.g., $\hat{v} = F(\tilde{v})$). In order to find this mapping, we use the residual learning strategy [17, 18, 19] where we predict the residual image (e.g., $\hat{v} - \tilde{v}$) from the noisy input image via a deep neural networks.

$$\hat{v} = F(\tilde{v}) = R(\tilde{v}) + \tilde{v}. \tag{9}$$

Note that the original mapping F will be closer to an identity mapping than the residual mapping R because noisy image is much more like the latent clean image than the residual image. Thus, the residual mapping can be easily trained for deeper neural networks while avoiding vanishing gradient issue.

Suppose we have noisy-clean image pairs $\{(\tilde{v}_i^{tr}, \hat{v}_i^{tr})\}_{i=1}^N$. Then, we minimize the mean squared error between the desired residual images and estimated ones from noisy input.

$$l(\mathbf{\Theta}) = \frac{1}{2N} \sum_{i=1}^{N} ||R(\tilde{v}_i^{tr}; \mathbf{\Theta}) - (\hat{v}_i^{tr} - \tilde{v}_i^{tr})||_F^2, \tag{10}$$

where $\mathbf{\Theta}$ represents the trainable weight parameters in deep neural networks.

The residual mapping $R(\cdot; \mathbf{\Theta})$ is defined as the layers of convolutional neural networks as shown in Fig. 1. First, we apply 64 filters of 3×3 convolution kernel to generate feature maps and then utilize rectified linear units (ReLU) [22] for neuron activation. Second, we apply 64 filters of $3\times3\times64$ convolution kernel for 15 layers. The batch normalization unit is added between convolution and ReLU to avoid internal covariate shift during mini-batch optimization [23]. Finally, we generate the residual image by applying 1 filter of $3\times3\times64$ convolution kernel.

By plugging the trained residual mapping $R(\cdot; \mathbf{\Theta})$ into eq. 5, 8 and 9, we can incorporate the deep residual learning based denoising operator into MBIR prior.

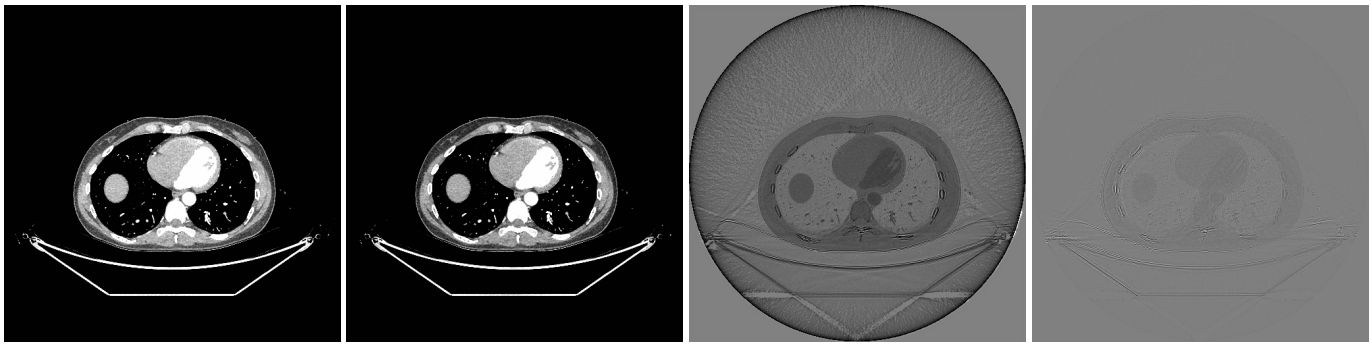
3. EXPERIMENTAL RESULTS

We validate our PnP MBIR with deep learning prior to the data acquired from a GE Revolution CT scanner with 50cm field of view. We collect two real patient scans of 256 slices with size of 512×512. We dedicate one patient scan to train the deep neural networks for image denoising and leave the other scan for reconstruction.

A deep neural network is trained to learn the relation between noisy FBP and clean ground-truth reconstructed from the same patient scan. For ground-truth, we generate the fully-converged MBIR image after 20 iterations. To create the training database, 40×40 patches are extracted in 256 image pairs with stride 20 and random permutations after data augmentation (e.g., horizontal and vertical flips, 90 degree rotation). We use the ADAM method [24] for optimization with gradually reduced learning rate from 10^{-3} to 10^{-5} with total 50 epochs. The size of mini-batch is set to 128. The training procedures are implemented using MatConvNet toolbox and take approximately 4 hours on a GTX TITAN X GPU.

Fig. 2 shows an example of the image denoising result via deep residual learning on the unseen testing scan. In Fig. 2(a), we observe that FBP is susceptible to noisy texture, particularly inside the soft tissue (e.g., cardiac muscle) when CT scan data is acquired at low dose. On the other hands, deep residual learning greatly reduces the noise and improves spatial resolution as depicted in Fig. 2(b). For reference, we also display the fully-converged ground-truth reconstruction after 20 iterations in Fig. 2(c). Then, our deep residual learning generates very close image to the fully converged ground-truth image in terms of texture inside the soft tissue. It is worth noting that computational time for image denoising based on the trained neural networks is very low under 10ms per slice.

With the trained deep neural networks for image denoising, we perform PnP MBIR on the unseen testing scan. For



(a) Standard MBIR (b) PnP MBIR (c) Ground-truth - Standard MBIR (d) Ground-truth - PnP MBIR

Fig. 3. Reconstruction results after 5 iterations: (a) Standard MBIR with MRF prior, (b) PnP MBIR with deep residual learning prior, (c) Subtracted image between ground-truth and standard MBIR, (d) Subtracted image between ground-truth and PnP MBIR. The display window of the intensity map is $[800, 1100]$ HU for the reconstructed images ((a) and (b)) and $[-100, 100]$ HU for the subtracted images ((c) and (d)), respectively. Note that both standard and PnP MBIR improve the image quality compared with FBP, while PnP MBIR produces closer image to the ground-truth image than standard MBIR.

comparison, we also generate the standard MBIR image with MRF prior [25]. Fig. 3 illustrates the reconstruction results for both MBIR methods after 5 iterations. As displayed in Fig. 3(a) and (b), the both methods improve the image quality in terms of noisy texture and spatial resolution compared with initial FBP image. Then, we calculate the subtracted images from the fully converged ground-truth to standard and PnP MBIR in Fig. 3(c) and (d), respectively. Our PnP MBIR shows the benefit in reducing background noise and DC offset over standard MBIR. This indicates that our PnP MBIR converges faster than standard MBIR as deep residual learning based prior quickly drives to the optimal solution.

For quantitative comparisons, we compute the root mean square difference (RMSD) between reconstruction results and the fully-converged ground-truth in whole 3D volume. Our PnP MBIR with deep residual learning prior significantly decreases the RMSD value from 24.35 HU to 5.99 HU compared with the standard MBIR with MRF prior, reflecting more accurate reconstruction.

4. CONCLUSION

In this paper, we present a novel MBIR with the prior model learned from a large dataset. Our method is based on the Plug-and-Play (PnP) framework which allows us to integrate the image denoising operator as a MBIR prior model. We use the residual learning for image denoising to train the deep convolutional neural networks which find the relation between the image pairs of noisy FBP and fully converged MBIR. Results on a real CT scan show that the deep residual learning is very effective in reducing the noise and enhancing the resolution in FBP. In addition, PnP MBIR with deep residual learning prior significantly improves the convergence speed compared with standard MBIR with MRF prior.

5. REFERENCES

[1] J.-B. Thibault, K. D. Sauer, C. A. Bouman, and J. Hsieh, "A three-dimensional statistical approach to improved image quality for multislice helical CT," *Medical Physics*, vol. 34, no. 11, pp. 4526–4544, Nov. 2007.

[2] R. Zhang, J-B Thibault, C. A. Bouman, K. D Sauer, and J. Hsieh, "Model-based iterative reconstruction for dual-energy X-ray CT using a joint quadratic likelihood model," *IEEE Trans. Med. Imag.*, vol. 33, no. 1, pp. 117–134, 2014.

[3] P. Jin, C. A. Bouman, and K. D. Sauer, "A model-based image reconstruction algorithm with simultaneous beam hardening correction for X-ray CT," *IEEE Trans. Comput. Imag.*, vol. 1, no. 3, pp. 200–216, 2015.

[4] J. Besag, "Spatial interaction and the statistical analysis of lattice systems," *Journal of the Royal Statistical Society. Series B (Methodological)*, pp. 192–236, 1974.

[5] I. A. Elbakri and J. A. Fessler, "Statistical image reconstruction for polyenergetic X-ray computed tomography," *IEEE Trans. Med. Imag.*, vol. 21, pp. 89–99, Feb. 2002.

[6] E. Y. Sidky and X. Pan, "Image reconstruction in circular cone-beam computed tomography by constrained, total-variation minimization," *Phys. Med. Biol.*, vol. 53, no. 17, pp. 4777–4807, 2008.

[7] R. Zhang, D. Ye, D. Pal, J.-B. Thibault, K. D. Sauer, and C. A. Bouman, "A Gaussian Mixture MRF for Model-Based Iterative Reconstruction," *IEEE Trans. Comput. Imag.*, vol. 2, no. 3, pp. 359–374, June 2016.

[8] S. V. Venkatakrishnan, C. A. Bouman, and B. Wohlberg, "Plug-and-play priors for model based reconstruction," in *IEEE Global Conf. Signal Inf. Process.*, 2013, pp. 945–948.

[9] S. Sreehari, S. V. Venkatakrishnan, B. Wohlberg, G. T. Buzzard, L. F. Drummy, J. P. Simmons, and C. A. Bouman, "Plug-and-play priors for bright field electron tomography and sparse interpolation," *IEEE Trans. Comput. Imag.*, vol. 2, no. 4, pp. 408–423, Dec. 2016.

[10] J. Eckstein and D. P. Bertsekas, "On the Douglas Rachford splitting method and the proximal point algorithm for maximal monotone operators," *Math. Program.*, vol. 55, no. 1-3, pp. 293–318, 1992.

[11] S. Boyd, N. Parikh, E. Chu, B. Peleato, and J. Eckstein, "Distributed optimization and statistical learning via the alternating direction method of multipliers," *Found. Trends Mach. Learn.*, vol. 3, no. 1, pp. 1–122, 2011.

[12] S. Zhang and E. Salari, "Image denoising using a neural network based non-linear filter in wavelet domain," in *Proc. IEEE Int. Conf. Acoust., Speech and Signal Process. (ICASSP)*, 2005.

[13] V. Jain and S. Seung, "Natural image denoising with convolutional networks," in *Proc. Neural Info. Process. Sys. (NIPS)*, 2009, pp. 769–776.

[14] P. Vincent, H. Larochelle, I. Lajoie, Y. Bengio, , and P. A. Manzagol, "Stacked denoising autoencoders: Learning useful representations in a deep network with a local denoising criterion," *Journal of Machine Learning Research*, vol. 11, pp. 3371–3408, Dec. 2010.

[15] H. C. Burger, C. J. Schuler, and S. Harmeling, "Image denoising: Can plain neural networks compete with BM3D?," in *Proc. IEEE Conf. Comput. Vision Pattern Recognition (CVPR)*, 2012, pp. 2392–2399.

[16] J. Xie, L. Xu, and E. Chen, "Image denoising and inpainting with deep neural networks," in *Proc. Neural Info. Process. Sys. (NIPS)*, 2012, pp. 341–349.

[17] K. He, X. Zhang, S. Ren, and J. Sun, "Deep Residual Learning for Image Recognition," in *Proc. IEEE Conf. Comput. Vision Pattern Recognition (CVPR)*, 2016.

[18] K. Zhang, W. Zuo, Y. Chen, D. Meng, and L. Zhang, "Beyond a Gaussian Denoiser: Residual Learning of Deep CNN for Image Denoising," *IEEE Trans. Imag. Process.*, vol. 26, no. 7, pp. 3142 – 3155, July 2017.

[19] E. Kang, J. Min, and J. C. Ye, "A deep convolutional neural network using directional wavelets for low-dose X-ray CT reconstruction," *Medical Physics*, vol. 44, no. 10, pp. e360–375, Oct. 2017.

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

[21] D. Kim and J. A. Fessler, "Optimized Momentum Steps for Accelerating X-ray CT Ordered Subsets Image Reconstruction," in *Int'l. Conf. on Image Formation in X-ray CT*, Jun. 2014, pp. 103–106.

[22] V. Nair and G. E. Hinton, "Rectied linear units improve restricted boltzmann machines," in *Proc. IEEE Int. Conf. Mach. Learn. (ICML)*, 2010, pp. 807–814.

[23] S. Ioe and C. Szegedy, "Batch normalization: Accelerating deep network training by reducing internal covariate shift," in *Proc. IEEE Int. Conf. Mach. Learn. (ICML)*, 2015, pp. 448–456.

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

[25] Z. Yu, L. Fu, D. Pal, J.-B. Thibault, C. A. Bouman, and K. D. Sauer, "Nested loop algorithm for parallel model based iterative reconstruction," in *International Conference on Fully 3D Image Reconstruction in Radiology and Nuclear Medicine*, 2013, pp. 197–200.