

# DEEP BACK PROJECTION FOR SPARSE-VIEW CT RECONSTRUCTION

Dong Hye Ye\*, Gregory T. Buzzard†, Max Ruby†, and Charles A. Bouman\*

\*School of Electrical and Computer Engineering, Purdue University, West Lafayette, IN, 47907

†Department of Mathematics, Purdue University, West Lafayette, IN, 47907

## ABSTRACT

Filtered back projection (FBP) is a classical method for image reconstruction from sinogram CT data. FBP is computationally efficient but produces lower quality reconstructions than more sophisticated iterative methods, particularly when the number of views is lower than the number required by the Nyquist rate. In this paper, we use a deep convolutional neural network (CNN) to produce high-quality reconstructions directly from sinogram data. A primary novelty of our approach is that we first back project each view separately to form a stack of back projections and then feed this stack as input into the convolutional neural network. These single-view back projections convert the encoding of sinogram data into the appropriate spatial location, which can then be leveraged by the spatial invariance of the CNN to learn the reconstruction effectively. We demonstrate the benefit of our CNN based back projection on simulated sparse-view CT data over classical FBP.

**Index Terms**— Deep Learning, Sparse-view CT, Image Reconstruction

## 1. INTRODUCTION

Computed Tomography (CT) is very important in applications ranging from health care and manufacturing to scientific exploration [1]. Traditional CT reconstruction algorithms require approximately  $n$  views, each with  $n$  channels, in order to reconstruct an  $n \times n$  image since this ensures Nyquist sampling. However, collecting  $n$  views is not always possible or practical. So for example, some security CT scanners use a fixed set of source positions, which leads to sparse views. In manufacturing applications, sparse view collection can reduce acquisition time, which reduces cost. In scientific applications, it is often not possible to collect a full set of views, particularly when imaging dynamically changing objects.

Regularized iterative methods, such as Model Based Iterative Reconstruction (MBIR), can form high-quality images from sparse views by incorporating into the reconstruction problem a forward model of the physics of the CT scanner together with a prior model of the object being imaged [2, 3,

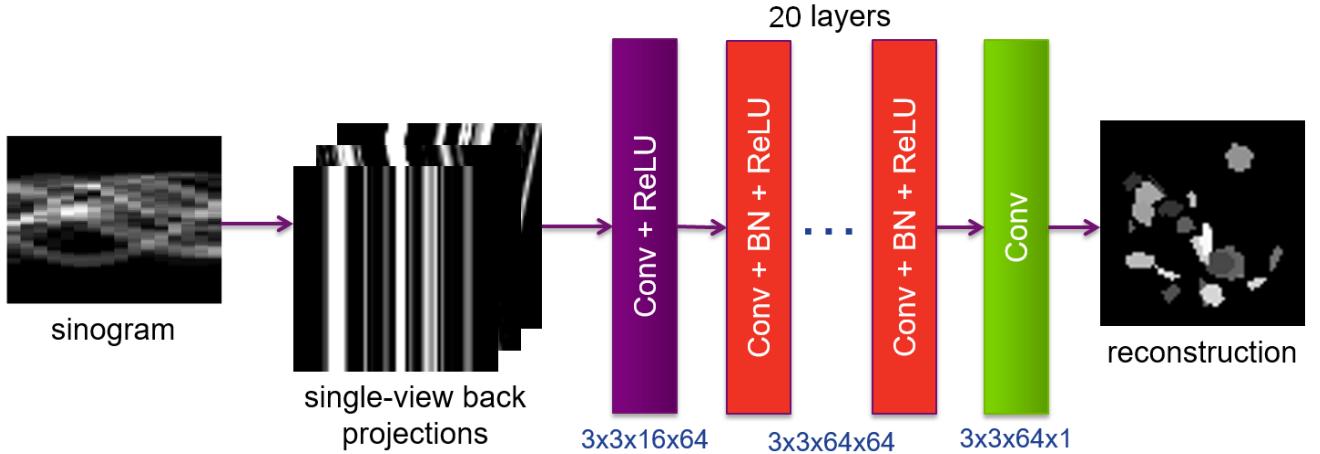
4]. However, while progress has been made in speeding up MBIR, it remains computationally expensive due to the nature of iterative optimization, which limits its use in practical applications.

Recently, there has been a great deal of interest in using deep convolutional neural networks (CNN) for image processing tasks. We make no attempt to survey this literature, but a few of many examples include effective uses of CNNs for denoising, image tracking, and object recognition [5, 6, 7]. More relevant for this paper, applications of CNNs for CT reconstruction have proliferated in recent years, with most methods focused on applying a CNN to a reconstructed image to reduce artifacts/increase quality. Some of these approaches use a CNN as a single-step denoiser, while others train a CNN to act as part of an iterative reconstruction method [8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19].

The power of CNNs is that they reduce the number of parameters of deep neural networks by imposing spatial invariance. The challenge of this in applying to sinograms is that the sinogram has information encoded in a spatially non-local way. In one approach to overcome this challenge, [20] use CNNs to learn approximate proximal maps for a version of the primal-dual algorithm. However, their approach uses a type of memory in both the primal and dual spaces and multiple learned proximal maps for different steps in the iteration, thus increasing the parameter space significantly. A related approach for MRI is given in [21]. In [22], fully connected layers are used to resolve the problem of spatial nonlocality of image reconstruction from sinogram data.

In this paper, we introduce a method called Deep Back Projection (DBP), in which we use a deep convolutional neural network to produce high-quality reconstructions directly from sinogram data. The challenge in this problem is finding a way to reorganize the sinogram data to make it amenable to processing by a CNN. In a primary novelty of our approach, we first create what we call a stacked back projection by back projecting each view separately to form an image of parallel lines of varying intensity (see Fig. 1). We then feed this stack as input into the convolutional neural network. This minimal, linear preprocessing step converts the nonlocal information encoded in the sinogram into spatially invariant information in reconstruction space, which can then be leveraged by the spatial invariance of the CNN to learn the full reconstruction

C.A. Bouman and G.T. Buzzard were partially supported by NSF CCF-1763896.



**Fig. 1.** Overview of Deep Back Projection (DBP): Given a sinogram, we generate single-view projections for all view angles. Multi-layer convolutional neural networks are trained to predict a reconstructed image. Note that we use batch normalization (BN) and rectified linear units (ReLU) for efficient training of deep neural networks.

effectively with relatively few parameters.

## 2. DEEP BACK PROJECTION

In Fig. 1, we show the overall framework of our Deep Back Projection (DBP) method. Given a sparsely measured CT sinogram (e.g., 16 view angles), we back project each view separately to form a stacked back projection. Then, we train deep convolutional neural networks to predict reconstructed image from this stack of single-view back projections. For the network architecture, we use a multi-layer convolutional filter with batch normalization and rectified linear units. By feeding single-view projections of unseen testing CT data into this trained network, we can reconstruct the image directly from the sinogram. In the following, we describe each component of our DBP in more detail.

### 2.1. The Stacked Back-Projection Tensor

Let  $y$  denote a set of 2D CT data with  $n$  channels and  $m$  views. Then  $y$  can be viewed as an  $n \times m$  sinogram image, where the  $i^{th}$  column corresponds to an individual view projection  $y_i$  taken at angle  $\theta_j$ . So then  $y = [y_1, \dots, y_m]$  where each element of the vector  $y_j \in \mathbb{R}^n$  is essentially the integral of the pixel intensities over a line with angle  $\theta_j$  and an offset from center that depends on the channel.

Each view projection,  $y_i$  is then back projected along the corresponding angle to form an  $n \times n$  image of parallel lines. Since this back projection operator is linear, the single-view back projected image  $Z_j \in \mathbb{R}^{n \times n}$  can be computed as

$$Z_j = B_j y_j, \quad (1)$$

where  $B_j$  is a single-view back projection operator at the

particular view angle  $\theta_j$ . The images  $Z_j \in \mathbb{R}^{n \times n}$  are then stacked to form a single tensor of dimension  $n \times n \times m$ .

$$\mathbf{Z} = [Z_1; \dots; Z_m]$$

We call  $\mathbf{Z}$  the back-projection tensor (see the example in Fig. 1 and 2).

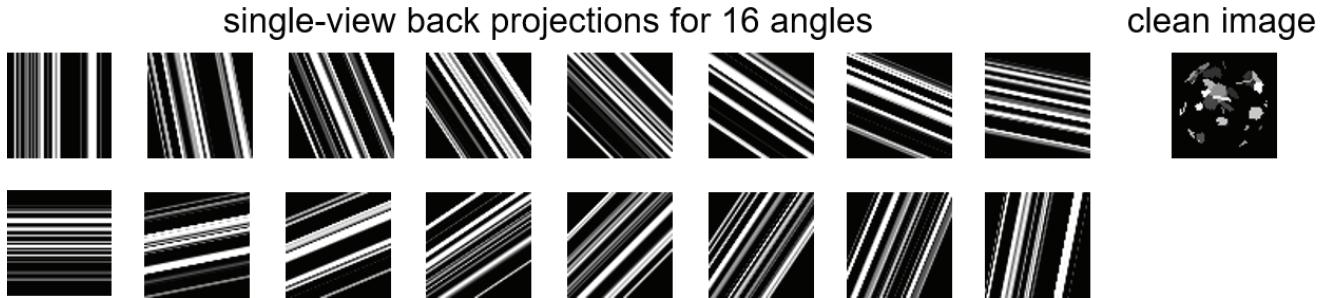
The key idea in our method is to use this back projection tensor,  $\mathbf{Z}$ , as the input tensor to a deep convolutional neural network. This approach converts a single sinogram into a stack of images each containing a collection of parallel lines of various intensities. Since each line is illustrated as constant intensity along its length, the information is stored in a highly redundant manner. However, the advantage is that the information is stored in a way that is spatially invariant, which is an important requirement in order for convolutional filters to be effective in a deep neural network.

### 2.2. Deep Learning for Image Reconstruction

The goal of reconstruction is to find the mapping  $F$  from the input sinogram  $y$  to the latent clean image  $x$  (e.g.,  $x = F(y)$ ). In order to find this mapping, we use a deep convolutional neural network [23] to predict the reconstructed image from the back projection tensor  $\mathbf{Z}$ .

Suppose we have  $K$  sinogram / clean image pairs denoted as  $\{(y_k^{tr}, x_k^{tr})\}_{k=1}^K$ . For each sinogram, we generate a stack of single-view back projections and build a training database  $\{(\mathbf{Z}_k^{tr}, x_k^{tr})\}_{k=1}^K$ . We then minimize the mean squared error between the desired clean images and estimated ones from sinogram.

$$l(\Theta) = \frac{1}{2K} \sum_{k=1}^K \|x_k^{tr} - F(\mathbf{Z}_k^{tr}; \Theta)\|_2^2, \quad (2)$$



**Fig. 2.** Example of dataset. Radon transform is applied to the clean synthetic image with 16 view angles to generate the CT sinogram. Single-view back projections from the sinogram are stacked to form the input for a convolutional neural network. Note that each single-view back projection shows lines with constant intensity with respect to the corresponding view angle.

where  $\Theta$  represents the trainable weight parameters in a deep neural network.

The reconstruction mapping  $F(\cdot; \Theta)$  is defined using the layers of a convolutional neural network as shown in Fig. 1. First, we apply 64 convolutional filter kernels of size  $3 \times 3 \times m$ , where  $m = 16$  for 16 view angles data, to generate feature maps. We then apply rectified linear units (ReLU) [24] for neuron activation. It is worth noting that we use 3-dimensional convolutional kernels to apply the convolution operator to the full stack of single-view back projections. Second, we apply 64 filters of  $3 \times 3 \times 64$  convolution kernel for 20 layers. A batch normalization unit is added between each convolution and a ReLU to avoid an internal covariate shift during mini-batch optimization [25]. Finally, we apply 1 filter of with a  $3 \times 3 \times 64$  convolution kernel to generate an image from the feature maps.

By feeding an unseen testing sinogram  $y^{te}$  into the trained reconstruction mapping  $F(\cdot; \Theta)$ , we can generate the reconstructed image  $\hat{x}^{te}$  directly from  $y^{te}$ .

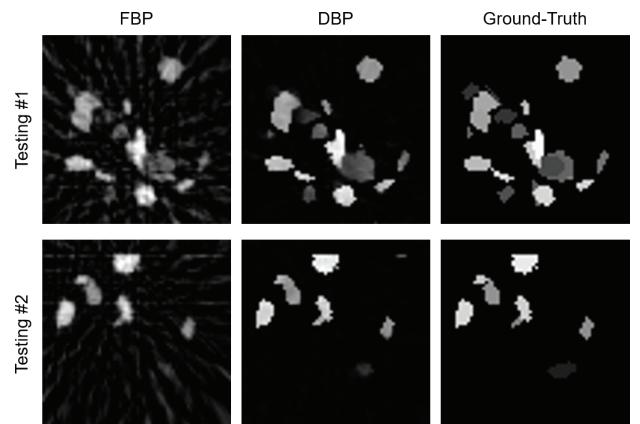
### 3. EXPERIMENTAL RESULTS

We validate our Deep Back Projection (DBP) algorithm using simulated data representing 16-view CT scans on multi-grain structures. We generate 100 synthetic noise-free images with size of  $64 \times 64$ . We then apply a Radon transform to the synthetic images to collect the sparse CT sinogram. We then form the back projection tensor for each sinogram. An example of this dataset is shown in Fig. 2. We dedicate 80 scans (subdivided into patches as described below) to train the deep neural network for image reconstruction and leave the other 20 scans for testing.

A deep neural network is trained to learn the relation between single-view back projections and the clean ground-truth image from the same scan. To create the training database, 256000 patches of size of  $8 \times 8$  are extracted from 80 stacked backprojection/clean image pairs using data augmentation (e.g., horizontal and vertical flips, 90 degree rotations).

It is worth noting that we extract patches from stacked back projections at the same location for all 16 view angles. We use the ADAM optimization [26] with gradually reduced learning rate from  $10^{-3}$  to  $10^{-5}$  with a total of 50 epochs. The size of a mini-batch is set to 128. The training procedures were implemented using the MatConvNet toolbox and took approximately 1 hour on a GTX TITAN X GPU.

Fig. 3 shows an example of the image reconstruction results on the unseen testing scans. For comparison, we first reconstruct the classical filtered back projection (FBP) images illustrated in the left column of Fig. 3. We observe that classical FBP is susceptible to spray artifacts in the entire image domain when the CT scan data is acquired with limited view angles. In contrast, our DBP algorithm greatly reduces the noise and improves spatial resolution as depicted in the middle column of Fig. 3. For reference, we also display the ground-truth image in the right column of Fig. 3. In terms of



**Fig. 3.** Example of reconstruction results on 2 testing scans. Classical FBP shows spray artifacts due to incomplete CT data on limited view angles. Our DBP reduces the noise and enhances the spatial resolution compared with classical FBP, matching the image quality of noise-free ground-truth.

**Table 1.** Quantitative Comparison between Reconstruction Results and Clean Ground-Truth Images

	PSNR (dB)	SSIM
FBP	$18.43 \pm 3.75$	$0.49 \pm 0.11$
DBP	$19.84 \pm 2.44$	$0.73 \pm 0.08$

texture, particularly inside the grain structure, the DBP reconstruction is very close to the noise-free ground-truth image. It is worth noting that the computational time for our DBP is under 10ms per slice, which is similar to that of classical FBP.

As a quantitative comparison, we report the peak signal to noise ratio (PSNR) and the structure similarity (SSIM) for 20 testing scans between reconstruction results and the clean ground-truth images in Table 1. Our DBP outperforms classical FBP with about 1.4 dB PSNR value increase reflecting significantly reduced noise. In addition, our DBP significantly increases the SSIM value from  $0.49 \pm 0.11$  to  $0.73 \pm 0.08$  compared with the classical FBP. This indicates that multi-grain structures are better reconstructed in our DBP with high spatial resolution than classical FBP.

#### 4. CONCLUSION

In this paper, we present a deep learning method for sparse-view CT reconstruction directly from the sinogram data, which we call deep back projection (DBP). Our method is based on the novel idea of constructing a “back-projection tensor” formed by a stack of single-view back projections. Importantly, the back-projection tensor contains all the information in the original sinogram, but it results in a structure with spatial invariance so that the convolutional filters in a deep convolutional neural network (CNN) can be most effective. We train a multi-layer CNN to find the relationship between the back projection tensor and the clean ground-truth images, and we then use this CNN to perform tomographic reconstruction. Results on a simulated 16-view CT scan show that our DBP method is very effective in reducing noise and enhancing the spatial resolution as compared to FBP. In addition, our DBP algorithm can preserve fine structures in the reconstructed image even with limited view angles, thus allowing faster acquisition without loss of quality.

#### 5. REFERENCES

[1] E. 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.

[2] J. Thibault, K. Sauer, J. Hsieh, and C. Bouman, “A three-dimensional statistical approach to improve image quality for multislice helical CT,” *Med. Phys.*, vol. 34, no. 11, pp. 4526–4544, 2007.

[3] 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.

[4] 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.

[5] O. Ronneberger, P. Fischer, and T. Brox, “U-Net: Convolutional Networks for Biomedical Image Segmentation,” in *Medical Image Computing and Computer-Assisted Intervention (MICCAI)*, 2015, vol. 9351, pp. 234–241.

[6] D. Ye, J. Li, Q. Chen, J. Wachs, and C. Bouman, “Deep Learning for Moving Object Detection from a Single Camera in UAVs,” in *IS&T International Symposium on Electronic Imaging*, 2018.

[7] 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.

[8] G. Wang, “A perspective on deep imaging,” *IEEE Access*, vol. 4, pp. 8914–8924, 2016.

[9] E. Kang, W. Chang, J. Yoo, and J. Ye, “Deep Convolutional Framelet Denoising for Low-Dose CT via Wavelet Residual Network,” *IEEE Trans. Med. Imag.*, vol. 37, pp. 1358 – 1369, 2018.

[10] D. Ye, S. Srivastava, J. Thibault, K. Sauer, and C. Bouman, “Deep Residual Learning for Model-Based Iterative CT Reconstruction using Plug-and-Play Framework,” in *Proc. IEEE Int. Conf. Acoust., Speech and Signal Process. (ICASSP)*, 2018.

[11] K. Batenburg and W. Kosters, “A Neural Network Approach to Real-Time Discrete Tomography,” in *International Workshop on Combinatorial Image Analysis*, 2006, pp. 389–403.

[12] E. Sidky, J. Jørgensen, and X. Pan, “Convex optimization problem prototyping for image reconstruction in computed tomography with the Chambolle-Pock algorithm,” *Phys. Med. Biol.*, vol. 57, pp. 3065–3091, 2012.

[13] K. Jin, M. McCann, E. Froustey, and M. Unser, “Deep Convolutional Neural Network for Inverse Problems in Imaging,” *IEEE Trans. Imag. Process.*, vol. 26, pp. 4509 – 4522, 2017.

[14] M. McCann, K. Jin, and M. Unser, “Convolutional Neural Networks for Inverse Problems in Imaging: A Review,” *IEEE Signal Processing Magazine*, vol. 34, no. 6, pp. 85 – 95, 2017.

[15] H. Gupta, K. Jin, H. Nguyen, M. McCann, and M. Unser, “CNN-Based Projection Gradient Descent for Consistent CT Image Reconstruction,” *IEEE Trans. Med. Imag.*, vol. 37, pp. 1440 – 1453, 2018.

[16] Y. Han and J.I. Ye, “Framing U-net via deep convolutional framelets: Application to sparse-view CT,” *IEEE Trans. Med. Imag.*, vol. 37, pp. 1418–1429, 2018.

[17] B. Kelly, T. Matthews, and M. Anastasio, “Deep Learning-Guided Image Reconstruction from Incomplete Data,” in *Neur. Info. Proc. Syst.*, 2017.

[18] H. Chen, Y. Zhang, M. Kalra, F. Lin, Y. Chen, P. Liao, J. Zhou, and G. Wang, “Low-Dose CT With a Residual Encoder-Decoder Convolutional Neural Network,” *IEEE Trans. Med. Imag.*, vol. 36, no. 12, pp. 2524–2535, Dec 2017.

[19] J. Schlemper, J. Caballero, J. V. Hajnal, A. N. Price, and D. Rueckert, “A deep cascade of convolutional neural networks for dynamic mr image reconstruction,” *IEEE Trans. Med. Imag.*, vol. 37, no. 2, pp. 491–503, Feb 2018.

[20] J. Adler and O. Öktem, “Learned Primal-dual Reconstruction,” *arXiv preprint arXiv:1707.06474*, 2017.

[21] Y. Yang, J. Sun, H. Li, and Z. Xu, “ADMM-Net: A Deep Learning Approach for Compressive Sensing MRI,” *ArXiv e-prints*, May 2017.

[22] Tobias Würfl, Florin C. Ghesu, Vincent Christlein, and Andreas Maier, “Deep learning computed tomography,” in *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2016*, Sébastien Ourselin, Leo Joskowicz, Mert R. Sabuncu, Gozde Unal, and William Wells, Eds., Cham, 2016, pp. 432–440, Springer International Publishing.

[23] 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.

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

[25] S. Ioffe 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.

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