Sparse-View CT Reconstruction using Recurrent Stacked Back Projection

Wenrui Li
Electrical and Computer Engineering
Purdue University
West Lafayette, IN, United States
li3120@purdue.edu

Gregery T. Buzzard
Mathematics
Purdue University
West Lafayette, IN, United States
buzzard@purdue.edu

Charles A. Bouman
Electrical and Computer Engineering
Purdue University
West Lafayette, IN, United States
bouman@purdue.edu

Abstract—Sparse-view CT reconstruction is important in a wide range of applications due to limitations on cost, acquisition time, or dosage. However, traditional direct reconstruction methods such as filtered back-projection (FBP) lead to low-quality reconstructions in the sub-Nyquist regime. In contrast, deep neural networks (DNNs) can produce high-quality reconstructions from sparse and noisy data, e.g. through post-processing of FBP reconstructions, as can model-based iterative reconstruction (MBIR), albeit at a higher computational cost.

In this paper, we introduce a direct-reconstruction DNN method called Recurrent Stacked Back Projection (RSBP) that uses sequentially-acquired backprojections of individual views as input to a recurrent convolutional LSTM network. The SBP structure maintains all information in the sinogram, while the recurrent processing exploits the correlations between adjacent views and produces an updated reconstruction after each new view. We train our network on simulated data and test on both simulated and real data and demonstrate that RSBP outperforms both DNN post-processing of FBP images and basic MBIR, with a lower computational cost than MBIR.

Index Terms—Sparse-view CT, CT reconstruction, deep learning, Recurrent Stacked Back Projection, LSTM

I. INTRODUCTION

In many applications of computed tomography (CT), the number of views that can be acquired is limited due to cost, acquisition time, hardware constraints, or allowable X-ray dosage. This motivates the challenging problem of reconstructing high quality images from sparse-view CT data. Traditional direct reconstruction algorithms such as filtered back projection (FBP) depend on the assumption of Nyquist sampling, hence can produce severe artifacts when given sparse-view data [1]. Iterative reconstruction algorithms such as model-based iterative reconstruction (MBIR) can dramatically improve the quality of sparse view reconstructions, but at the cost of higher computation [2]–[4].

More recently, the use of deep neural networks (DNNs) has emerged as a fundamentally new approach to tomographic reconstruction [5] with the advantages that a) it can greatly

reduce computation, and b) given sufficient training data, it can be trained to incorporate complex prior information [6]. One way to use DNNs for CT reconstruction is to preprocess sinogram data [7]–[9] and then apply a standard reconstruction; however, this approach does not fully utilize the capabilities of DNNs in the image domain. A second approach uses DNNs for post-processing of an image reconstructed using a simple method such as FBP [10]–[12]; this approach leverages the capabilities of DNNs in the image domain but does not have full access to all sinogram data. Finally, direct reconstruction from the sinogram domain using DNNs as in [6] is very memory and computationally expensive [13].

In this paper, we introduce a direct-reconstruction DNN method called Recurrent Stacked Back Projection (RSBP) that uses single-view backprojections as sequential input to a recurrent convolutional LSTM network, whose output is then post-processed by a U-Net. The Stacked Back Projection (SBP) maintains all information in the sinogram [14], while the recurrent processing exploits the correlations between adjacent views and produces an updated reconstruction after each new view. We train our network on simulated data and test on both simulated and real data and demonstrate that RSBP outperforms similar DNN post-processing of FBP images or SBP tensors and also outperforms basic MBIR, with a lower computational cost than MBIR.

II. RECURRENT STACKED BACK PROJECTION

In Fig. 1, we illustrate the network architecture of our RSBP method, in which individual backprojections are stacked in the image domain and fed sequentially to a convolutional LSTM, whose image output is then fed to a U-Net for final reconstruction. In section II-A, we introduce the SBP tensor, which contains all information from the sparse-view sinogram. In section II-B, we introduce each component of our proposed network in more detail.

A. Stacked Back Projection Tensor

We consider a CT reconstruction problem in which an unknown image $x \in \mathbb{R}^{N \times N}$ is observed via N detectors and M views to yield noisy measurements $y \in \mathbb{R}^{N \times M}$. Each column vector $y_j \in \mathbb{R}^N$ approximates the integral of attenuation along the line with angle θ_j according to the model

$$y_j = A_j x + w_j, \tag{1}$$
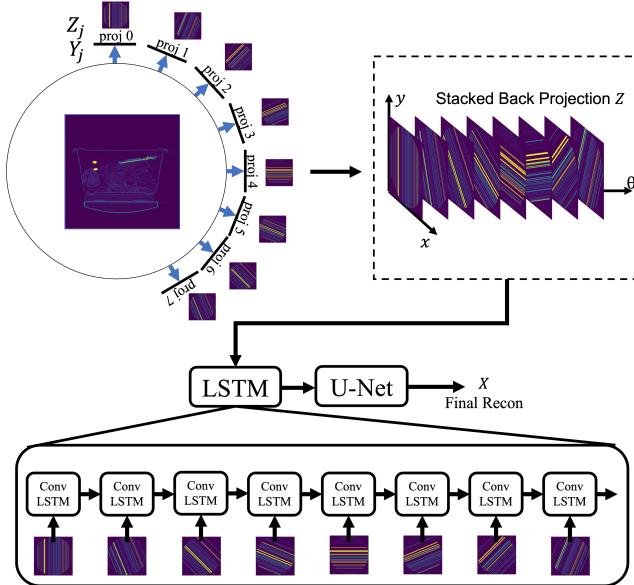
Fig. 1: **Conceptual view of RSBP architecture:** a) The single-view projections Z_j are back-projected from a single projection y_j, then collected into a tensor called the Stacked Back Projection. b) Each single-view back-projection is input to the LSTM sequentially, then further processed by a U-Net.

where A_j is the Radon transform operator with angle $\theta_j = \frac{j\pi}{M}$ for $j = \{0, 1, .., M-1\}$, and $w_j \in \mathbb{R}^N$ represents signal-dependent noise by $w_j \sim N(0, \mathrm{diag}(\frac{1}{\lambda_0} \exp(A_j x)))$, where λ_0 is the empty scan photon count.

To obtain the SBP tensor, we first apply single-view FBP to each measured projection to obtain

$$Z_j = A_j^\dagger(y_j), \tag{2}$$

where $A_j^\dagger$ is the inverse Radon transform using a single-view FBP operator with angle θ_j. As in [14], the images $Z_j \in \mathbb{R}^{N \times N}$ are then stacked to form a rank 3 SBP tensor, in our case with dimensions $N \times N \times M$:

$$\boldsymbol{Z} = \{Z_1; Z_2; ...; Z_M\}. \tag{3}$$

In contrast to an FBP image, the SBP tensor contains all information from sparse-view CT measurements. But like the FBP, the SBP converts the complex spatial correlations in the sinogram back to the image domain, which allows for the use of efficient processing by convolutional neural networks (CNNs), as shown in Fig. 2(b).

B. Convolutional LSTM Network with SBP

Two problems with the SBP as used in [14] are that it requires advance knowledge of the number views and requires that all the single-view backprojections be stored in memory. That is, post-processing the SBP with a CNN as in Fig. 2(b) treats each single-view back projection as an independent channel, like RGB channels in a color image, and processes these channels simultaneously.
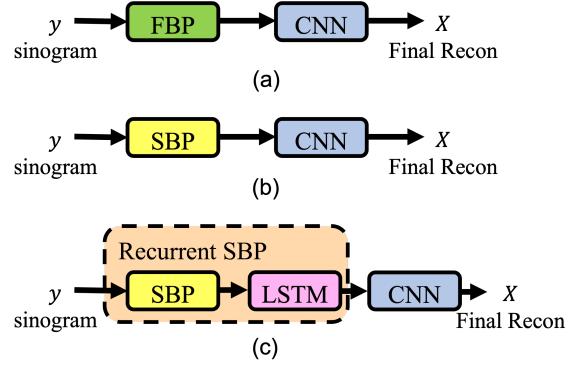
Fig. 2: **Evolution towards RSBP:** a) CNN post-processing of FBP image; b) simultaneous CNN post-processing of SBP tensor; c) sequential input of SBP backprojections to convolutional LSTM followed by CNN postprocessing.
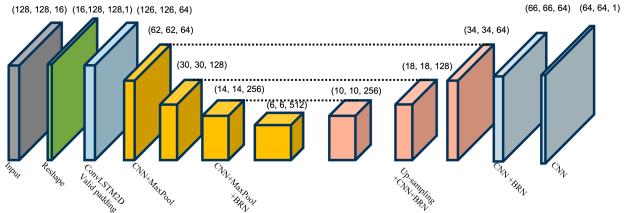
Fig. 3: **RSBP architecture:** The input SBP is fed sequentially to a convolutional LSTM, with output then processed by a U-Net using valid padding. With valid padding, the output image is smaller than the input image, so we use only the central region of the corresponding ground truth image to calculate loss functions and quality metrics.

To address these problems, we introduce a convolutional LSTM [15] to process these single-view backprojections sequentially. Most applications of recurrent networks such as the LSTM involve time-dependent measurements, such as speech processing, while most CT reconstruction methods use all available views simultaneously. However, as shown in Fig. 1, the SBP not only contains all single-view backprojections but also is sequentially acquired. This sequential acquisition process and the two problems identified above lead naturally to propose a recurrent network. The advantage of a convolutional LSTM in this context is that it has the spatial invariance of a CNN and can be trained with sequentially acquired data. We call this approach recurrent stacked back-projection (RSBP), as shown in Fig. 2(c).

C. Network architectures

The full RSBP-CNN architecture is illustrated in Fig. 3, which shows the initial processing by the LSTM followed by a U-Net [16]. We first reshape the SBP tensor from (Rows, Columns, Views) to (Views, Rows, Columns, Channels) and then feed this to the convolutional LSTM layer, with kernel size 3×3 for both the input transformations and recurrent transformations. From the LSTM we obtain a

Projection Geometry:	Parallel beam geometry
Number of Views:	8 or 16 equi-spaced angles in $[0, \pi]$
Pixel Pitch p:	0.186 cm
FOV:	47.616 cm
Water X-ray density μ:	0.17 cm^{-1} ($\sim$100 keV)
Photon dosage per proj (λ_0):	1600

(Rows, Columns, Channels) tensor and pass it to a U-Net. We use batch renormalization in most of the downsampling and upsampling blocks for better convergence and stability of the network [17], [18]. In the final layer, we use convolution with a 3×3 kernel to generate a single output image.

III. IMPLEMENTATION

We implement and train RSBP-CNN using Keras with a loss function described below. We use previously reconstructed full-view CT data to serve as ground truth; we use 8 or 16 simulated projection views to serve as input to RSBP. We evaluate the method on both simulated sinograms and real sinograms.

A. Dataset Generation

We generate training and testing data using the ALERT Task Order 4 dataset [19], which consists of 188 full-view sinograms and high-quality 3D reconstructions of packed suitcases using data from an Imatron Scanner. First, we split the data into 153 3D reconstructions for training and 35 for testing. From each reconstruction, we randomly select 28 slices to yield 4284 slices for training and 980 slices for testing. Next, we generate simulated projections using the scaling and noise modeling approach described in [20]. For each view angle θ_j we use the forward model

$$y_j = (\mu p) A_j \left(\frac{x}{1000} \right) + w_j \; ; \qquad (4)$$

here division by 1000 scales the image from (modified) Hounsfield units[1] to units in which water is 1 and air is 0, A_j is the raw radon transform for θ_j, μ and p are the water X-ray density and pixel pitch, respectively, and $w_j \sim N(0, R)$ with R a diagonal matrix with entries $\frac{\exp\{A_j x\}}{\lambda_0}$ to model measurement noise. The physical constants are defined in Table I.

From the scaled simulated sinogram y, we apply single-view back projection and inverse scaling, then stack as in Fig. 1 to generate the SBP. This process results in pairs of ground truth images and corresponding simulated SBP tensors in Hounsfield units. The SBP tensor is scaled back to units in which water is 1 before sending to the NN, then scaled back to Hounsfield units before evaluating the loss function.

B. Loss Function

For reconstructing an X-ray CT image for security applications, the range from 0 to 2000 Hounsfield units (HU) (air=0

[1]We use modified Hounsfield units in which air is 0 and water is 1000.

HU, water=1000 HU) is critical. Therefore, we use a modified MSE loss to train our models:

$$L_{MSE} = \|f(x_{true}) - f(\hat{x})\|_2^2, \qquad (5)$$

where x_{true} is the ground truth (cropped to the valid region as described in Fig. 3), $\hat{x}$ is the reconstructed image, $f(x) = \frac{x}{|x|+2000}$, and both x_{true} and $\hat{x}$ are in Hounsfield units.

C. Training Process

We implemented all NN models in Tensorflow Keras and used one NVIDIA Tesla V100 GPU for training. In order to reduce memory requirements and increase batch size during training [21], we used randomly cropped patches of size 128×128 from the SBP tensors (original image size 320×320) as input and calculated the loss between the output (64×64) and the center of the corresponding patches from the ground truth images. In inference mode, we applied the trained model on full image slices. We used the Adam optimizer with learning rate = 0.0002. For each model, we trained for 120 epochs and shuffled the training dataset after each epoch.

IV. EXPERIMENTS

We compare our proposed RSBP-CNN model with several alternatives, including (a) MBIR [22] using the QGGMRF prior, (b) FBP-CNN, and (c) SBP-CNN. FBP-CNN and SBP-CNN, illustrated in Fig. 2, both use the architecture in Fig. 3 with a convolutional layer in place of the LSTM; we train them with the same strategy use for RSBP.

We examine reconstructions using either 8 views or 16 views. In each of these cases we trained each of FBP-CNN, SBP-CNN, and RSBP-CNN for a total of 6 trained networks. We tested these DNN methods and MBIR on simulated data as described above and on real sinogram data of 9 different 3D objects from the ALERT Task Order 3 dataset [23].

Since we apply the raw inverse radon transform on real sinogram data, we apply inverse scaling to the SBP as before. The parameters are the same as in Table I except that now μ is 0.159 cm^{-1} ($\sim$130 keV).

In each reconstruction, we computed NRMSE (normalized root mean square error) between reconstruction and ground truth using

$$NRMSE = \frac{\|x_{true} - \hat{x}\|_2}{\|x_{true}\|_2} \; . \qquad (6)$$

For results using real sinogram data, we use an MBIR reconstruction with 720 views as ground truth.

Fig. 4 shows reconstructions with each of the 4 methods using each of 8-view and 16-view sinogram data. The images and NRMSE show that RSBP-CNN improves over SBP-CNN, which improves over FBP-CNN, in both cases for both 8-view and 16-view sinogram data. RSBP-CNN eliminates most of the streak artifacts seen in the MBIR and FBP-CNN reconstructions and preserves more details than SBP-CNN.

The improvement from using SBP is consistent with results of [14] and the fact that the SBP includes all sinogram data,

Proposed Method

8 Views

Ground Truth

16 Views

|			
MBIR NRMSE: 0.5517	FBP-CNN NRMSE: 0.4968	SBP-CNN NRMSE: 0.4703	RSBP-CNN **NRMSE: 0.4460**
MBIR NRMSE: 0.4782	FBP-CNN NRMSE: 0.4022	SBP-CNN NRMSE: 0.3646	RSBP-CNN **NRMSE: 0.3529**
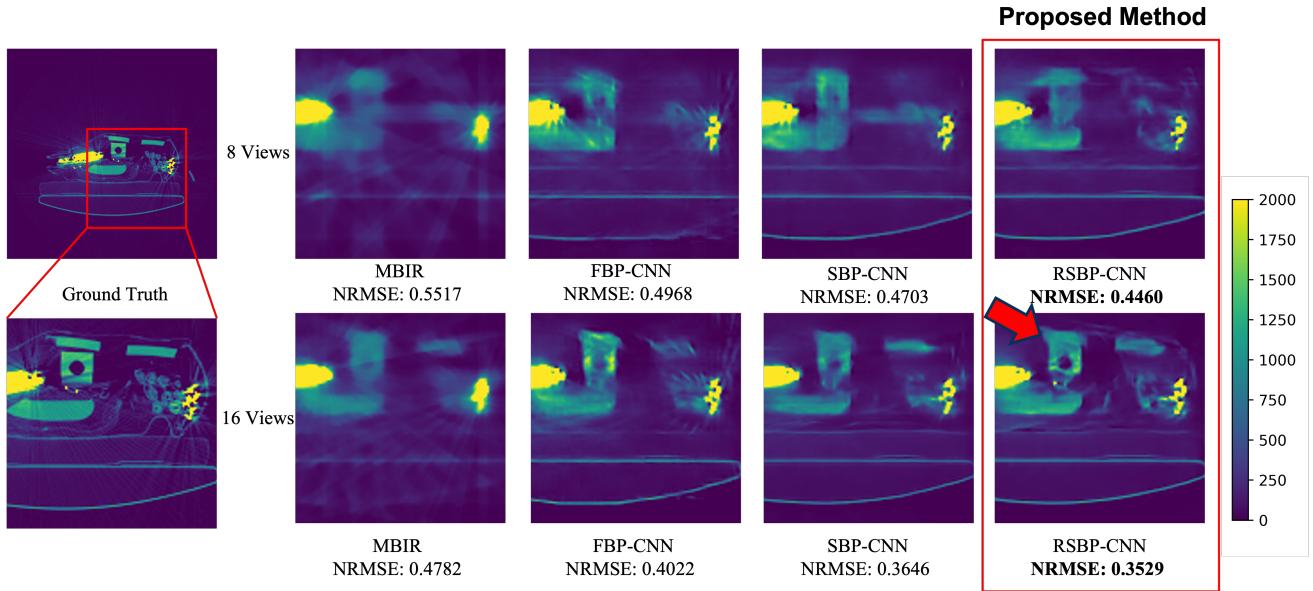
Fig. 4: **Comparisons of multiple reconstruction algorithms with simulated sinogram:** *Left 2 panels:* Ground truth image used to generate simulated sinogram data and a cropped detail used for display comparisons. *Right panels:* Reconstructions using MBIR, FBP-CNN, SBP-CNN and RSBP-CNN. The top row uses 8 views and the bottom row uses 16 views. The NRMSE is for the full image rather than the displayed cropped image. The display range is from 0 (air) to 2000 Hounsfield units (HU). The RSBP-CNN reconstruction has more detail and lower NRMSE relative to the other methods.

which is not true of the FBP. The improvement from SBP-CNN to RSBP-CNN is more subtle; in principle SBP-CNN has all the information need to perform as well as RSBP-CNN. We hypothesize that sequential processing of individual backprojections promotes better use of correlations between adjacent views; in addition, there may also be interactions between our training scheme and the reduced memory requirements of RSBP relative to simultaneous processing of the SBP.

Fig. 5 repeats the experiment of Fig. 4 using real sinogram data. In this case we use a full-view MBIR reconstruction as ground truth to compute the NRMSE. Once again we see that RSBP improves over SBP, which improves over FBP.

In Table II, we see that the performance gains from using RSBP and SBP are consistent across multiple trials. That table shows the mean and standard deviation of NRMSE over the simulated test data (980 scans) and the real data (900 scans). In all cases, there is a signifcant drop in NRMSE when going from FBP-CNN to SBP-CNN and again from SBP-CNN to RSBP-CNN. Hence RSBP shows real promise for truly sparse-view reconstruction.

V. Conclusion and Future work

In this paper, we introduced a novel, direct, deep learning algorithm RSBP-CNN that combines SBP, LSTM, and U-Net to reconstruct CT images from 8 or 16 view sinogram data. Our method couples the representational efficiency of the SBP as input with the computational and inferential efficiency of a convolutional LSTM. The use of an LSTM brings the ability to learn embedded order information from the sequentially-acquired SBP, thus enabling better reconstruction than the same network with a CNN in the first layer; LSTM also yields a reduction in memory requirements due to the sequential processing of back projections in the SBP. Our proposed RSBP-CNN outperforms other methods in terms of visual and metric comparisons on both simulated and real sinogram data. Future work will investigate modifications to promote the ability of a single RSBP-CNN to produce reconstructions without specifying in advance the number of views.

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," *Medical physics*, vol. 35, no. 2, pp. 660–663, 2008.

[2] S. J. Kisner, E. Haneda, C. A. Bouman, S. Skatter, M. Kourinny, and S. Bedford, "Model-Based CT Reconstruction from Sparse Views," in *Second International Conference on Image Formation in X-Ray Computed Tomography*, 2012, pp. 444–447.

[3] 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, 2007.

[4] S. V. Venkatakrishnan, C. A. Bouman, and B. Wohlberg, "Plug-and-Play Priors for Model Based Reconstruction," in *2013 IEEE Global Conference on Signal and Information Processing*, 2013, pp. 945–948.

[5] K. H. Jin, M. T. McCann, E. Froustey, and M. Unser, "Deep Convolutional Neural Network for Inverse Problems in Imaging," *IEEE Transactions on Image Processing*, vol. 26, no. 9, pp. 4509–4522, 2017.

[6] S. Ravishankar, J. C. Ye, and J. A. Fessler, "Image Reconstruction: From Sparsity to Data-Adaptive Methods and Machine Learning," *Proceedings of the IEEE*, vol. 108, no. 1, pp. 86–109, 2020.

[7] H. Lee, J. Lee, and S. Cho, "View-Interpolation of Sparsely Sampled Sinogram using Convolutional Neural Network," in *Medical Imaging 2017: Image Processing*, vol. 10133. International Society for Optics and Photonics, 2017, p. 1013328.

TABLE II: Mean / standard deviation of NRMSE for different reconstruction algorithms on multiple datasets

Method		MBIR	FBP-CNN	SBP-CNN	RSBP-CNN
Simulated sinogram	8-View	0.567 / 0.084	0.478 / 0.084	0.454 / 0.081	**0.428 / 0.084**
	16-View	0.498 / 0.080	0.381 / 0.076	0.375 / 0.070	**0.335 / 0.070**
Real sinogram	8-View	0.553 / 0.088	0.590 / 0.112	0.519 / 0.085	**0.506 / 0.092**
	16-View	0.462 / 0.087	0.411 / 0.089	0.405 / 0.077	**0.368 / 0.085**
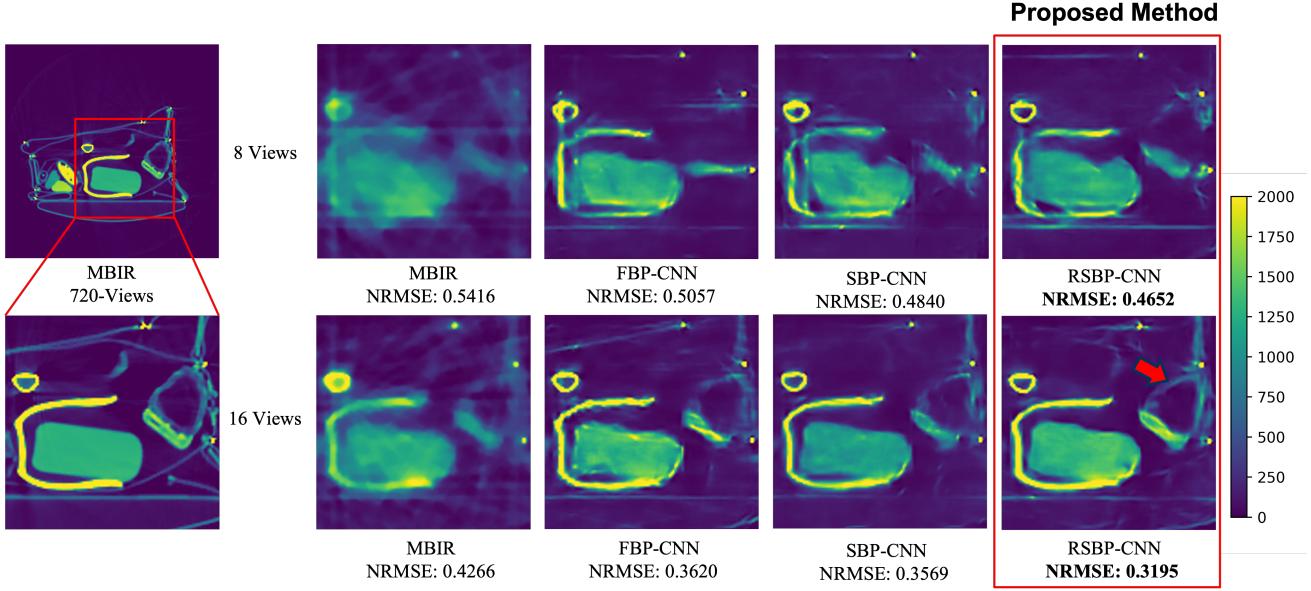
Fig. 5: **Comparisons of multiple reconstruction algorithms with real sinogram:** *Left 2 panels:* Ground truth image generated from full view real sinogram data using MBIR and a cropped detail used for display comparisons. *Right panels:* Reconstructions using MBIR, FBP-CNN, SBP-CNN, and RSBP-CNN. The top row uses 8 views and the bottom row uses 16 views. The NRMSE is for the full image rather than the displayed cropped image. The display range is from 0 (air) to 2000 Hounsfield units (HU). The RSBP-CNN reconstruction has clearer edges and lower NRMSE relative to the other methods.

[8] H. Lee, J. Lee, H. Kim, B. Cho, and S. Cho, "Deep-Neural-Network Based Sinogram Synthesis for Sparse-View CT Image Reconstruction," *IEEE Transactions on Radiation and Plasma Medical Sciences*, vol. 3, no. 2, pp. 109–119, Mar. 2019.

[9] M. U. Ghani and W. C. Karl, "Deep Learning-Based Sinogram Completion for Low-Dose CT," in *2018 IEEE 13th Image, Video, and Multidimensional Signal Processing Workshop (IVMSP)*. IEEE, 2018, pp. 1–5.

[10] A. Ziabari, D. H. Ye, S. Srivastava, K. D. Sauer, J.-B. Thibault, and C. A. Bouman, "2.5D Deep Learning for CT Image Reconstruction Using A Multi-GPU Implementation," in *52nd Asilomar Conference on Signals, Systems and Computers*, Oct. 2018, pp. 2044–2049.

[11] Y. Han and J. C. Ye, "Framing U-Net via Deep Convolutional Framelets: Application to Sparse-View CT," *IEEE transactions on medical imaging*, vol. 37, no. 6, pp. 1418–1429, 2018.

[12] Z. Zhang, X. Liang, X. Dong, Y. Xie, and G. Cao, "A Sparse-View CT Reconstruction Method based on Combination of DenseNet and Deconvolution," *IEEE transactions on medical imaging*, vol. 37, no. 6, pp. 1407–1417, 2018.

[13] B. Zhu, J. Z. Liu, S. F. Cauley, B. R. Rosen, and M. S. Rosen, "Image Reconstruction by Domain-Transform Manifold Learning," *Nature*, vol. 555, no. 7697, pp. 487–492, Mar. 2018.

[14] D. H. Ye, G. T. Buzzard, M. Ruby, and C. A. Bouman, "Deep Back Projection for Sparse-View CT Reconstruction," in *2018 IEEE Global Conference on Signal and Information Processing (GlobalSIP)*. IEEE, 2018, pp. 1–5.

[15] X. Shi, Z. Chen, H. Wang, D. Y. Yeung, W. K. Wong, and W. C. Woo, "Convolutional LSTM Network: A Machine Learning Approach for Precipitation Nowcasting," *Advances in neural information processing systems*, vol. 2015, pp. 802–810, 2015.

[16] O. Ronneberger, P. Fischer, and T. Brox, "U-Net: Convolutional Networks for Biomedical Image Segmentation," in *International Conference on Medical image computing and computer-assisted intervention*. Springer, 2015, pp. 234–241.

[17] S. Ioffe and C. Szegedy, "Batch normalization: Accelerating deep network training by reducing internal covariate shift," in *International conference on machine learning*. PMLR, 2015, pp. 448–456.

[18] S. Ioffe, "Batch renormalization: Towards reducing minibatch dependence in batch-normalized models," *arXiv preprint arXiv:1702.03275*, 2017.

[19] C. Crawford, "Advances in automatic target recognition (atr) for ct based object detection system—final report," in *Task order HSHQDC–12–J–00429*. Dept. Homeland Security Center Excellence, ALERT, Boston, MA, USA, 2014.

[20] G. T. Buzzard and C. A. Bouman, "Ct-tutorial jupyter notebook," Jul 2019. [Online]. Available: https://github.com/gbuzzard/CT-Tutorial/blob/master/CT_Tutorial.ipynb

[21] H. Lee, J. Lee, H. Kim, B. Cho, and S. Cho, "Deep-neural-network-based sinogram synthesis for sparse-view ct image reconstruction," *IEEE Transactions on Radiation and Plasma Medical Sciences*, vol. 3, no. 2, pp. 109–119, 2018.

[22] X. Wang, A. Sabne, S. Kisner, A. Raghunathan, C. Bouman, and S. Midkiff, "High performance model based image reconstruction," *ACM SIGPLAN Notices*, vol. 51, no. 8, pp. 1–12, 2016.

[23] C. Crawford, H. Martz, and W. C. Karl, "Research and development of reconstruction advances in ct-based object detection systems – final report," in *Tech. Rep. HSHQDC-12-J-00056*. Dept. Homeland Security Center Excellence, ALERT, Boston, MA, USA, 2013.