Comparison of deep learning approaches to low dose CT using low intensity and sparse view data

Thomas Humphries*, Dong Si, Sean Coulter, Matthew Simms, and Ruiwen Xing

School of STEM, University of Washington Bothell, Bothell WA 98011.

ABSTRACT

Recently there has been considerable interest in using deep learning to improve the quality of low dose CT (LDCT) images. LDCT may be achieved by reducing the beam intensity, or by acquiring sparse-view data at full beam intensity. Additionally, if reducing beam intensity, one can consider denoising either the raw (sinogram) data, or the reconstructed image. We compare the performance of a convolutional neural network (CNN) in improving image quality using three approaches: denoising low-intensity images, denoising low-intensity sinograms prior to reconstruction, and denoising sparse-view images. Our results indicate that images produced from low-intensity data are superior to images produced from sparse-view data, after correction by the CNN. Additionally, in the low-intensity case, denoising in the sinogram or image domain provides comparable image quality.

Keywords: computed tomography, deep learning, low dose, denoising, sparse view

1. INTRODUCTION

Concerns about the radiation dose associated with computed tomography (CT) has motivated significant research interest in low dose CT (LDCT) in recent years. The most straightforward way of achieving LDCT is to reduce the intensity of the X-ray beam used in a normal dose (NDCT) scan, by lowering the tube current. Since X-ray measurements are stochastic, however, reducing beam intensity produces noisier data, thereby lowering image quality. A second approach is to maintain the same beam intensity and acquire fewer projections over the same acquisition arc. This approach, known as sparse view CT, produces images which suffer from streaking artifacts, due to the undersampling in the projection domain. An example of simulated NDCT, reduced intensity and sparse view images which illustrate these artifacts is provided in Figure 1.

A number of approaches have been proposed for improving image quality in both the reduced intensity and sparse view cases. These include filtering or interpolating the raw (sinogram) data; reconstructing images iteratively using regularizers to penalize undesirable features, or applying post-reconstruction artifact reduction

---

*Corresponding Author. E-mail: `thumphri@uw.edu`

Figure 1. Simulated NDCT (left), reduced intensity (middle) and sparse view (right) images. Reduced intensity image is simulated by reducing beam intensity by a factor of 15; sparse view by reducing number of views by a factor of 15. Details on simulation parameters are provided in Section 2.

techniques. Over the last two to three years, there has also been a dramatic increase in interest in the application of techniques from deep learning[1,2] to address these problems. Briefly, deep learning refers to a class of machine learning methods which use deep artificial neural networks to recognize patterns in data at multiple levels of representation. In a typical setup, the neural network is trained on a large set of input data to produce a desired output; for example, one may provide the network with a series of images and ask it to output a classification (does the image represent a dog, cat, bird?) After training, the network is then applied to new datasets to see if it produces correct outputs.

The power of deep learning lies in the multi-level approach used to extract and identify important features of the input. The field has exploded in recent years due in large part to the seminal 2012 paper of Krizhevsky et al,[3] whose deep convolutional neural network (CNN) was shown to significantly outperform other state-of-the-art methods on image classification tasks. Typically, a CNN consists of several different types of component:

1. Convolutional layers apply a number of filters to the image to extract important feature. A filter can be viewed as an operation that acts on every pixel of the image by replacing its value with a weighted average of other pixel values in its neighborhood. The weights of the filter are not known beforehand, and are learned by the CNN during training.

2. Pooling layers reduce the size of the input by merging pixel values within a neighborhood into a single value; for example, by taking the maximum value in the neighborhood. This operation increases the effective field of view of subsequent filters and allows the CNN to identify larger-scale features of the image.

3. Activation layers consist of a simple function applied to the data between convolutional or pooling layers. A popular choice is the rectified linear unit (ReLU), $f(x) = \max(x, 0)$, which sets any negative values to zero, leaving other values unchanged. The role of the activation function is to introduce nonlinearity to the CNN, allowing it to model a much wider range of functions.

To train a CNN, a loss function is first defined which measures the discrepancy between the actual and desired outputs of the CNN for the training dataset. An optimized set of weights for the convolutional layers is then determined using an iterative optimization algorithm. As the number of weights is typically on the order of hundreds of millions, graphical processing unit (GPU) based computing is necessary to accelerate the computation. Following training, the CNN is then applied to a separate test dataset to assess its performance.

In the context of CT imaging, the input is typically an image whose quality is compromised due to one or more of the factors listed above, and the desired output is a higher quality image. A number of approaches and applications of deep learning to CT imaging are highlighted in a recent special issue of IEEE Transactions on Medical Imaging (see Ref.,[4] along with cited papers). Some examples of applications to CT imaging include the use of neural networks to denoise low-intensity LDCT images,[5–9] improve quality of images reconstructed from sparse-view and limited-angle data,[10–14] and remove artifacts caused by metal or beam hardening.[15–17] In general, deep learning can be applied to problems in CT in several ways. Some approaches reconstruct the image using filtered back projection, with a CNN behaving as a pre or post-processing step. Other approaches use the CNN in tandem with an iterative reconstruction method; for example, to learn an appropriate regularizer,[9,11] or to learn a projection onto a set of feasible images that can be applied in every iteration.[18] Additionally, a number of different network architectures are possible, including so-called U-nets[10,13] and generative adversarial networks (GAN).[7,8]

The present work aims to address two questions relevant to the application of deep learning to LDCT:

1. Is a CNN more effective at improving the quality of an LDCT image achieved by reducing beam intensity, or an image produced from sparse view data corresponding to an equivalent dose reduction?

2. If beam intensity is reduced, is it more effective to train a CNN to denoise the image reconstructed from the LDCT sinogram, or to denoise the LDCT sinogram and then reconstruct the image?

The first question is of interest since either technique can be used to reduce dose. Reducing the beam intensity results in lower-quality measurements which nonetheless fully sample the projection domain, while sparse view imaging preserves the statistical quality of each projection, but undersamples the projection space. Thus, there

is a trade-off between the two approaches. Previous work in the context of total variation (TV)-regularized reconstruction has indicated that that acquiring many low-quality projections is preferable to acquiring few high-quality projections.[19] We investigate whether the same is true in the context of a CNN-based approach.

Our rationale for the second question is as follows. Most work on denoising using CNNs has been performed in the image domain, either during or post reconstruction. In the image domain, however, the noise in a given patch of the image is nonlocal; for example, streaks may be caused by rays passing through materials with high attenuation coefficients that are not included in the patch. Thus, two patches containing the same type of tissue may experience different noise patterns depending on their location in the image. In the sinogram domain, however, a measurement m_ℓ along the line ℓ is given by

$$m_\ell = -\log \frac{\hat{I}}{I_0}, \text{ where}$$

$$\hat{I} = \text{Poisson}\left(I_0 \exp\left\{-\int_\ell \mu(x)\, dx\right\}\right)$$

Here I_0 is the inital beam intensity, and the expression in braces represents the line integral along ℓ through the attenuation function $\mu(x)$. Thus, the level of noise in m_ℓ is directly correlated with the value of the measurement; the larger m_ℓ, the greater the noise. We therefore hypothesize that it may be easier for a CNN to learn the mapping from LDCT to NDCT data in the sinogram domain. Additionaly, sinogram correction is commonly applied in CT to account for beam hardening and metal artifacts, so a sinogram-based denoising approach may be useful in cases where these issues are also present.

2. METHODS

2.1 Data generation

Clinical contrast-enhanced CT images were obtained from The Cancer Imaging Archive's QIN LUNG CT dataset.[20] Seventy slices from a single patient study were read in and converted from Hounsfield Units to linear attenuation coefficients at a reference energy of 70 keV. We then used the Michigan Image Reconstruction Toolbox (MIRT)[21] to simulate monoenergetic fan-beam data. We simulated 900 views over $360°$, with source-object and source-detector distances of 80 cm and a fan angle of $35°$. Poisson noise was added to the pre-log data, proportionally to a simulated initial beam intensity, I_0. In total, we generated the following data sets:

1. NDCT data: 900 views with $I_0 = 10^6$.
2. low intensity (LI) data: 900 views with $I_0 = 2.0 \times 10^5$, 1.0×10^5, and 6.67×10^4.
3. sparse view (SV) data: 180, 90, and 60 equally spaced views with $I_0 = 10^6$.

The low intensity and sparse view data sets were generated so as to result in equivalent dose reductions relative to the NDCT data set (one fifth, one tenth, and one-fifteenth the original dose, respectively). Images were reconstructed from these datasets using Matlab's `ifanbeam` command, which implements filtered backprojection for fan-beam data. Through this process, we generated the following paired training datasets for the CNN:

1. NDCT image and low intensity image (denoted by LI-image),
2. NDCT sinogram and low intensity sinogram, (LI-sino) and
3. NDCT image and sparse view image (SV-image).

Each of these paired datsets was generated for the three dose levels described above, resulting in a total of nine datasets overall.
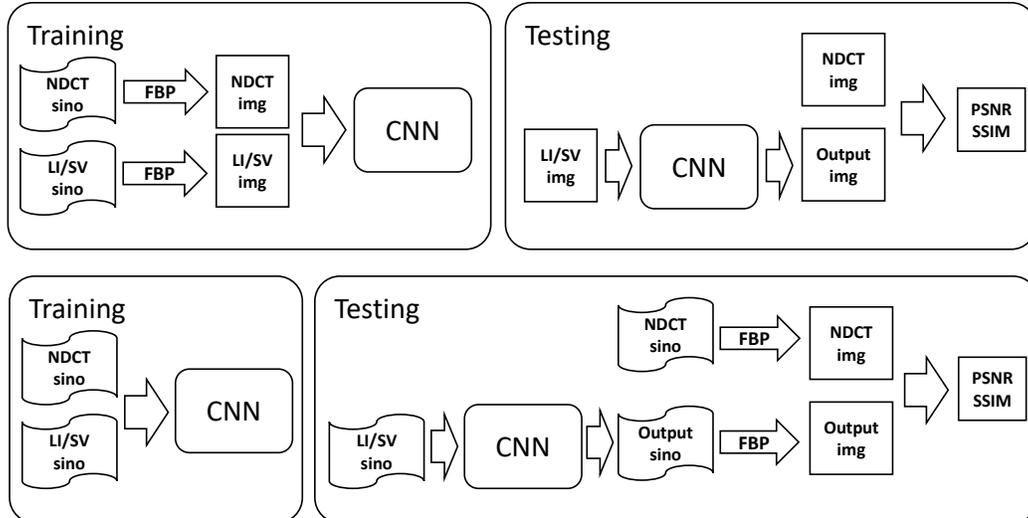
Figure 2. Flowchart indicating training and testing procedures. Top: For the LI-image and SV-image datasets, the CNN is trained to minimize ℓ_2 loss between reconstructed images. In the test phase, output from the CNN is compared with the NDCT image. Bottom: For the LI-sino dataset, the CNN is trained to minimize ℓ_2 loss between sinograms. In the test phase, the output from the CNN is used to reconstruct a denoised image, which is compared with the NDCT image.

2.2 CNN architecture and training

Our CNN is implemented in Tensorflow and based on the DnCNN of 22. While the original DnCNN[*] is trained to denoise photographs which have had Gaussian noise added to them, it can be straightforwardly adapted to CT imaging with pairs of reconstructed images (NDCT and LDCT) as input. DnCNN consists of an initial layer of sixty-four 3×3 convolutions with rectified linear unit (ReLU) activation; fifteen layers with sixty-four filters of size $3 \times 3 \times 64$, using ReLU and batch normalization; and a final convolutional layer with one $3 \times 3 \times 64$ filter to recover the output. No pooling layers are included. Given a normal dose image or sinogram, y, and associated low dose image or sinogram z, the DnCNN attempts to learn the residual mapping $\mathcal{R}(z)$ such that $y \approx z - \mathcal{R}(z)$. The network is trained using ℓ_2 loss.

The same fifty CT image slices (or corresponding sinograms) from the patient study were used to train each network, while the remaining twenty were used as test data. Both testing and training data sets contained slices corresponding to the pelvic, abdominal, and thoracic regions of the patient. The network was trained on small overlapping patches extracted from each training image, allowing us to generate thousands of training data from a single CT image slice. For the LI-image and SV-image datasets, 14,400 patches of size 32×32 were extracted from each image, resulting in a total of 720,000 training patches. For the LI-sino dataset, patches were of size 36×27 to account for the different dimension of the sinogram (900 angles by 729 bins). All networks were trained over 50 epochs using Tensorflow's ADAM optimizer.

In testing, the output from the CNN was compared directly with the NDCT images for the LI-image and SV-image datasets. For the LI-sino dataset, the output from the CNN (a denoised sinogram) was used to reconstruct an image using `ifanbeam`, which was then compared with the corresponding NDCT image. Image quality was measured using peak signal-to-noise ratio (PSNR) and structural similarity (SSIM).[23] The training and testing process for these networks is shown in Figure 2.

3. RESULTS

The average PSNR and SSIM values of the twenty test images are displayed in Table 1. It is apparent that, prior to correction, the SV images are of much lower quality than the LI images. While all networks substantially improve the quality of the input image, we observe that the quality of the SV images is still significantly poorer

---

[*] <https://github.com/crisb-DUT/DnCNN-tensorflow>

Table 1. Average PSNR and SSIM values of the 20 test images. First three columns are values for the uncorrected low intensity image and images corrected in the image and sinogram domain, respectively; last two columns are for the uncorrected and corrected sparse view images.

		LI	LI-image	LI-sino	SV	SV-image
$\frac{1}{5}$ dose	PSNR	41.96	46.83	46.62	33.34	42.05
	SSIM	0.9240	0.9756	0.9734	0.7317	0.9612
$\frac{1}{10}$ dose	PSNR	39.35	46.05	45.74	27.11	36.53
	SSIM	0.8802	0.9739	0.9709	0.5087	0.9177
$\frac{1}{15}$ dose	PSNR	37.69	45.52	45.12	24.10	33.20
	SSIM	0.8447	0.9725	0.9689	0.4002	0.8770

than that of the LI images; this is especially true as the number of views is reduced to 60 (one-fifteenth the original dose). This finding agrees with the previously published results for TV-regularized iterative reconstruction.[19] Additionally, the quality of the LI denoised images is fairly consistent across all three dose levels, while the quality of the corrected sparse view images diminishes every time the dose (i.e. number of views) is reduced. The average PSNR and SSIM values for the LI-image and LI-sino approaches indicate that the performance of both methods is comparable. It is interesting that even though the LI-sino approach is trained to minimize ℓ_2 error in the sinogram domain, it results in comparable image quality measured in the image domain.

The output of the trained CNNs for one of the twenty test images is shown in Figure 3. The quality of the LI-image and LI-sino denoised images is essentially the same when taking into account the entire image. For this particular test image, the SV-image reconstruction is also of comparable quality at the dose level of one-fifth, though the quality decreases as the dose is further reduced. Interestingly, when zooming in on the two regions of interest in blue and red, the computed SSIM values within the ROI suggest that the LI-sino reconstruction is of somewhat higher quality than the LI-image reconstruction. Qualitatively, the LI-sino reconstruction appears somewhat smoother than the LI-image reconstruction as well. When investigating further by examining other small regions of interest in this and the other test images, however, we did not observe a consistent difference in performance between the two methods.

4. CONCLUSIONS AND FUTURE WORK

We have studied the performance of a simple CNN-based approach to LDCT using both low intensity and sparse view scans. Our results indicate that the CNN is better able to correct LDCT images achieved via low intensity scans, versus those obtained from sparse view scans. It is possible that one reason for the discrepancy in performance is that the network did not include any pooling layers, which increase the effective field of view of the filters being applied by convolutional layers. While both sparse-view artifacts and image noise in CT are spatially varying, it seems likely that sparse-view artifacts vary more significantly across the entire image. Thus, a network architecture including pooling, such as a U-net,[10, 13] may be more effective.

In addition, we found that our network achieved comparable image quality by denoising a low intensity sinogram prior to reconstruction, versus denoising the corresponding image after reconstruction. This may be of interest in cases where the data are noisy and also contaminated by effects such as beam hardening, which are often addressed by sinogram-domain corrections. While the data simulated in our experiments were monoenergetic, it would be interesting to test the method on polyenergetic low-intensity data to see if there is any advantage to a sinogram-domain correction.

REFERENCES

[1] LeCun, Y., Bengio, Y., and Hinton, G., "Deep learning," *Nature* **521**(7553), 436 (2015).
[2] Goodfellow, I., Bengio, Y., and Courville, A., [*Deep Learning*], MIT press Cambridge (2016).
[3] Krizhevsky, A., Sutskever, I., and Hinton, G. E., "Imagenet classification with deep convolutional neural networks," in [*Advances in neural information processing systems*], 1097–1105 (2012).
[4] Wang, G., Ye, J. C., Mueller, K., and Fessler, J. A., "Image reconstruction is a new frontier of machine learning," *IEEE Trans. Med. Imag.* **37**(6), 1289–1296 (2018).
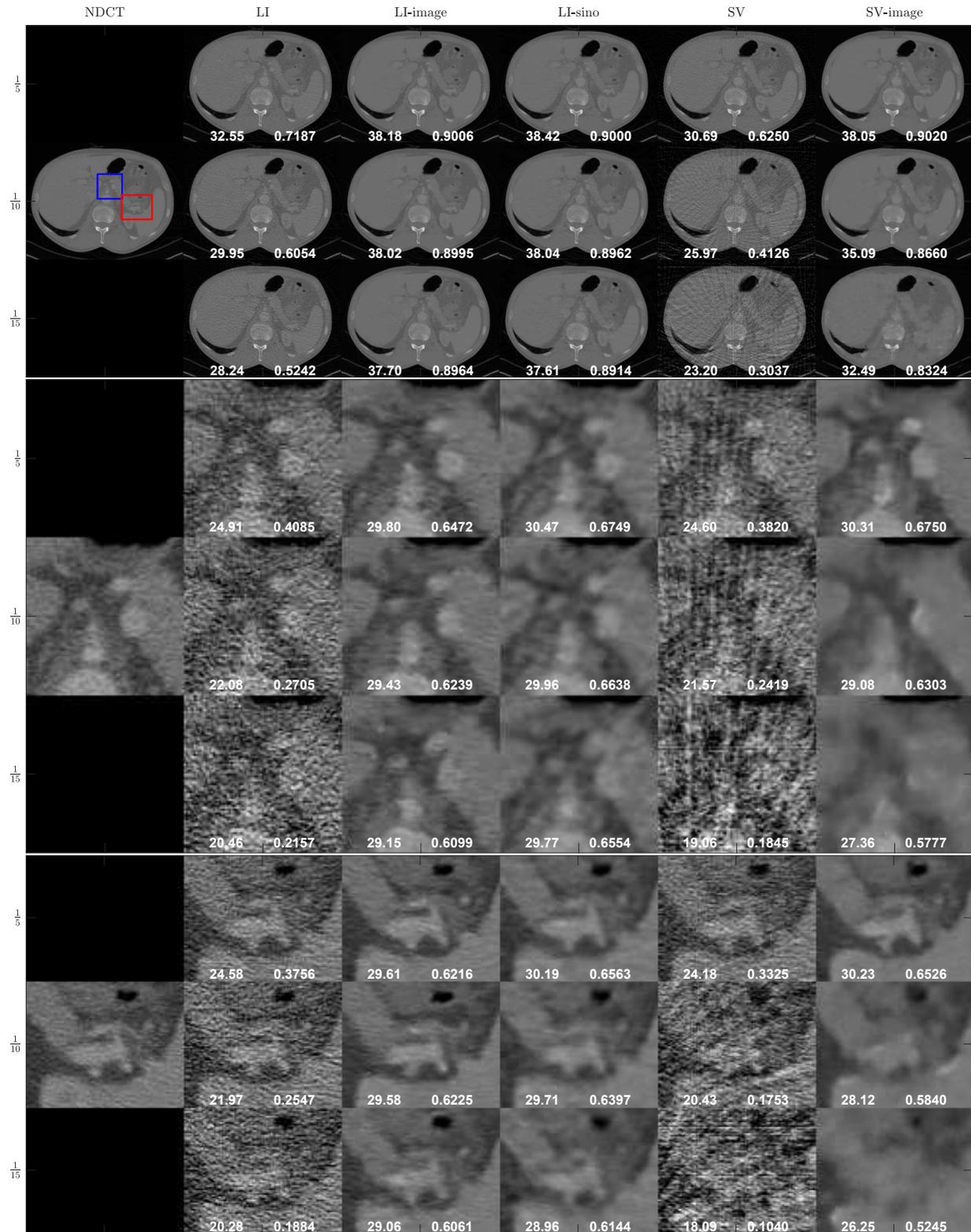
Figure 3. Test images corrected using different CNN approaches. First column contains reference NDCT image; remaining five columns follow the same order as Table 1. Rows correspond to different dose levels; top three rows are for entire image, middle three rows for zoomed in region in blue, bottom three for region in red. White numbers give the PSNR and SSIM, respectively, computed for that image or region.

[5] Kang, E., Min, J., and Ye, J. C., "A deep convolutional neural network using directional wavelets for low-dose x-ray CT reconstruction," *Medical physics* **44**(10) (2017).

[6] Chen, H., Zhang, Y., Kalra, M. K., Lin, F., Chen, Y., Liao, P., Zhou, J., and Wang, G., "Low-dose CT with a residual encoder-decoder convolutional neural network," *IEEE Trans. Med. Imag.* **36**(12), 2524–2535 (2017).

[7] Wolterink, J. M., Leiner, T., Viergever, M. A., and Išgum, I., "Generative adversarial networks for noise reduction in low-dose CT," *IEEE Trans. Med. Imag.* **36**(12), 2536–2545 (2017).

[8] Yang, Q., Yan, P., Zhang, Y., Yu, H., Shi, Y., Mou, X., Kalra, M. K., Zhang, Y., Sun, L., and Wang, G., "Low dose CT image denoising using a generative adversarial network with wasserstein distance and perceptual loss," *IEEE Trans. Med. Imag.* (2018).

[9] Zheng, X., Ravishankar, S., Long, Y., and Fessler, J. A., "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).

[10] Zhang, Z., Liang, X., Dong, X., Xie, Y., and Cao, G., "A sparse-view CT reconstruction method based on combination of DenseNet and deconvolution.," *IEEE Trans. Med. Imag.* **37**(6), 1407–1417 (2018).

[11] Chen, H., Zhang, Y., Chen, Y., Zhang, J., Zhang, W., Sun, H., Lv, Y., Liao, P., Zhou, J., and Wang, G., "LEARN: Learned experts assessment-based reconstruction network for sparse-data CT," *IEEE Trans. Med. Imag.* (2018).

[12] Würfl, T., Hoffmann, M., Christlein, V., Breininger, K., Huang, Y., Unberath, M., and Maier, A. K., "Deep learning computed tomography: Learning projection-domain weights from image domain in limited angle problems," *IEEE Trans. Med. Imag.* **37**(6), 1454–1463 (2018).

[13] Han, Y. and Ye, J. C., "Framing u-net via deep convolutional framelets: Application to sparse-view CT," *IEEE Trans. Med. Imag.* **37**(6), 1418–1429 (2018).

[14] Gu, J. and Ye, J. C., "Multi-scale wavelet domain residual learning for limited-angle CT reconstruction," *arXiv preprint arXiv:1703.01382* (2017).

[15] Zhang, Y. and Yu, H., "Convolutional neural network based metal artifact reduction in x-ray computed tomography," *IEEE Trans. Med. Imag.* **37**(6), 1370–1381 (2018).

[16] Ghani, M. U. and Karl, W. C., "Deep learning based sinogram correction for metal artifact reduction," *Electronic Imaging* **2018**(15), 472–1 (2018).

[17] Cong, W. and Wang, G., "Monochromatic CT image reconstruction from current-integrating data via deep learning," *arXiv preprint arXiv:1710.03784* (2017).

[18] Gupta, H., Jin, K. H., Nguyen, H. Q., McCann, M. T., and Unser, M., "CNN-based projected gradient descent for consistent CT image reconstruction," *IEEE Trans. Med. Imag.* **37**(6), 1440–1453 (2018).

[19] Jørgensen, J. S. et al., "Effect of sparsity and exposure on total variation regularized X-ray tomography from few projections," *Proc. 4th Int. Mtg. on Image Formation in X-ray CT* , 279–82 (2016).

[20] Clark, K. et al., "The cancer imaging archive (TCIA): maintaining and operating a public information repository," *Journal of digital imaging* **26**(6), 1045–1057 (2013).

[21] Fessler, J., "Michigan image reconstruction toolbox," (2017).

[22] Zhang, K. et al., "Beyond a Gaussian denoiser: Residual learning of deep CNN for image denoising," *IEEE Transactions on Image Processing* **26**(7), 3142–3155 (2017).

[23] Wang, Z., Bovik, A. C., Sheikh, H. R., and Simoncelli, E. P., "Image quality assessment: from error visibility to structural similarity," *IEEE Transactions on Image Processing* **13**, 600–612 (April 2004).