2021 43rd Annual International Conference of the
IEEE Engineering in Medicine & Biology Society (EMBC)
Oct 31 - Nov 4, 2021. Virtual Conference

Self-supervised Projection Denoising for Low-Dose Cone-Beam CT

Kihwan Choi[1]

Abstract— **We consider the problem of denoising low-dose x-ray projections for cone-beam CT, where x-ray measurements are typically modeled as signal corrupted by Poisson noise. Since each projection view is a 2D image, we regard the low-dose projection views as examples to train a convolutional neural network. For self-supervised training without ground truth, we partially blind noisy projections and train the denoising model to recover the blind spots of projection views. From the projection views denoised by the learned model, we can reconstruct a high-quality 3D volume with a reconstruction algorithm such as the standard filtered backprojection. Through a series of phantom experiments, our self-supervised denoising approach simultaneously reduces noise level and restores structural information in cone-beam CT images.**

I. INTRODUCTION

Onboard cone-beam computed tomography (CBCT) is commonly used for patient setup and adaptive replanning in radiation therapy. Although CBCT is capable of quickly producing on-treatment patient anatomy, the repeated utilization of CBCT has increased concern about the risk of the radiation dose [1]. To reduce radiation exposure of patients, low-dose protocols with tube current modulation and lower tube voltage have been applied to CBCT imaging [2], [3]. However, low-dose CBCT protocols also decrease signal-to-noise ratio (SNR) in x-ray measurements, which may adversely affect their clinical usefulness [4], [5].

Deep learning has become a dominant machine learning tool in visual recognition and image processing [6]–[8]. Such advances in deep learning are being used to denoise low-dose CT (LDCT) images. The majority of previous deep learning approaches for LDCT denoising depend on supervised learning with the normal-dose CT (NDCT) image as a ground truth of LDCT images [9]–[12]. Although the supervised learning approaches have shown outstanding denoising performance, acquiring LDCT-NDCT pairs requires redundant CT scans of the same patients, which deliver additional radiation dose [9]–[12].

Recent studies on self-supervised learning have demonstrated that denoising networks can be trained without the use of clean references [13]–[16]. A pioneering work known as noise-to-noise training [13] has shown that training a denoising network from pairs of noisy images, which are independent noisy samples of the same underlying ground truth, is equivalent to learning to predict the clean image. As a step further, self-supervised learning, which requires neither clean targets nor noisy image pairs in denoising tasks, has been proposed [14]–[16]. The studies on self-supervised denoising rely on blinding some pixels of the input image and training CNNs to predict the blinded pixels from other pixels. When the noise is independent across pixels and the images are structured, the self-supervised learning approaches in fact predict the true images.

In this paper, we focus on denoising low-dose CBCT projection views by training a convolutional neural network without ground truth. As an alternative to supervised learning, we propose a self-supervised learning approach with blind filtering, which learns structures from low-dose x-ray projections alone. With the projection views denoised by the self-supervised learning model, we can reconstruct a 3D CBCT object with the standard filtered-backprojection (FBP) algorithm. The experimental phantom studies show that the self-supervised denoising network simultaneously reduces noise level and restores anatomical information in CBCT images as well as projections.

II. METHOD

A. Statistical Loss for CBCT Projection Denoising

If we vectorize the attenuation coefficients of a 3D CBCT object as $\mathbf{u} \in \mathbb{R}^n$, the discrete version of line integral can be expressed as

$$p_i = \mathbf{a}_i^T \mathbf{u}, \tag{1}$$

where a_{ij} is the length of intersection between the i-th path and j-th voxel u_j. With the incident photon number N_{0i}, the measured photon number y_i has Poisson distribution

$$y_i \sim Poisson(N_{0i} e^{-p_i}), \tag{2}$$

where the mean is given by $\bar{y}_i = N_{0i} e^{-p_i}$. By collecting the line integrals which belong to the same projection view, we denote the vectorization of 2D projection view by $\mathbf{y} \in \mathbb{R}^m$.

In order to obtain the line integral p_i for $i = 1, \ldots, m$, our goal is to train a convolutional neural network $G : \mathbb{R}^n \to \mathbb{R}^n$ which estimates $\bar{\mathbf{y}}$ from the measured projection view $\mathbf{y}$. For the measured projection view $\mathbf{y}$, we can define residual sum of squares (RSS) as

$$RSS = \|\bar{\mathbf{y}} - \mathbf{y}\|_{\ell_2}, \tag{3}$$

where $\bar{\mathbf{y}}$ is the mean of projection view $\mathbf{y}$. With the given data set $Y = \{\mathbf{y}^{(1)}, \ldots, \mathbf{y}^{(K)}\}$ consisting of K projection views, we define our loss function

$$L_{\mathrm{RSS}}(G; Y) = \mathbb{E}_Y \|G(\mathbf{y}) - \mathbf{y}\|_{\ell_2}, \tag{4}$$

where $\mathbb{E}_Y$ denotes the empirical expectation over $\mathbf{y} \in Y$. Here, we expect that the denoised projection views are close to the mean of projection views, *i.e.*, $G(\mathbf{y}) \approx \bar{\mathbf{y}}$.

*This work was supported by Korea Institute of Science and Technology (KIST) Institutional Program (Project No. 2E31122).

[1]Kihwan Choi is with Center for Bionics, Korea Institute of Science and Technology (KIST), Seoul, Korea (e-mail: kihwanc@kist.re.kr).

B. Self-supervised Projection Denoising

Optimizing G to minimize the loss function L_{RSS} in (4) results in a trivial solution since the identical transform, which holds $G(\mathbf{y}) = \mathbf{y}$ for any $\mathbf{y} \in \mathbb{R}^n$, achieves the global minimum. Alternatively, we apply self-supervised learning to the loss function L_{RSS} in (4). Following the theoretical framework [14], we employ blind filtering for self-supervised learning.

We first define a set of blind spots $J \subset \{1, \ldots, n\}$, which is a subset of projection dimension, and $\mathbf{y}_J \in \mathbb{R}^{|J|}$ denotes a subset of projection view $\mathbf{y} \in \mathbb{R}^n$ corresponding to J. If a function $f(\mathbf{y})$ does not depend on the value of $\mathbf{y}_J$, we call f a J-invariant function with respect to J. We use a moving average filter $f_J : \mathbb{R}^n \to \mathbb{R}^n$, of which kernel value at the center is assigned to zero [14], [15] for blind filtering. By striding over the input projection data, the filter blinds the central values and interpolates the values with the neighboring values within the projection.

Regarding a set $\mathcal{J}$ as a collection of blind spots, we propose the self-supervision loss as the following:

$$L_{\text{self}}(G; Y) = \sum_{J \in \mathcal{J}} \mathbb{E}_Y \| G(f_J(\mathbf{y}))_J - \mathbf{y}_J \|_{\ell_2}, \tag{5}$$

which is the empirical expectation of the self-similarity over the LDCT projection views $\mathbf{y}$ in the training set Y.

C. Image Reconstruction

With the denoising network G, the estimated projection data $\hat{\mathbf{y}}$ can be written as:

$$\hat{\mathbf{y}} = G(\mathbf{y}), \tag{6}$$

where $\mathbf{y}$ is the measured projection data. Using (2), the estimated line integral can be written as:

$$\hat{p}_i = -\ln\left(\frac{\hat{y}_i}{N_{0i}}\right), \tag{7}$$

from which we can collect the estimated projection views. Finally, we can estimate 3D object $\hat{\mathbf{u}}$ by solving

$$\hat{\mathbf{p}} = \mathbf{A}\hat{\mathbf{u}}, \tag{8}$$

where $\mathbf{A}$ and $\hat{\mathbf{p}}$ are the system matrix and the collection of estimated line integrals $\hat{p}_i$ to reconstruct $\mathbf{u}$. For the single circular CBCT scan, we can reconstruct $\hat{\mathbf{u}}$ from the estimated projections with the standard filtered backprojection such as FDK algorithm [5].

D. Network Architecture

As the baseline of our network architecture, we adapt a contemporary denoising network (DnCNN) which has shown significant noise reduction performance [17]. From the perspective of network architecture, DnCNN extends VGG network [7] and learns residuals between the input and target pairs. For our self-supervised learning framework, we replace the simple convolution layers of DnCNN with residual blocks, which have been used in training image generative networks to transfer image style using a small number of instances [18]–[20]. The denoising network is fully convolutional and capable of handling full-size images without pre- and post-processing.

III. EXPERIMENTS AND RESULTS

A. Data Acquisition

The experimental CBCT projection data of anthropomorphic phantom were acquired by using an Acuity system (Varian Medical Systems). The number of projection views for a full 360° rotation is 680 and the total time for the acquisition about 1 min. The dimension of each acquired projection image is 397×298 mm^2, containing 1024×768 pixels.

During the projection data acquisition, the x-ray tube current was set at 80 mA, and the duration of the x-ray pulse at each projection view was 10 ms. The tube voltage was set to 125 kVp. The projection data were acquired in full-fan mode in a single circular scan with a bowtie filter. The distances of source-to-axis and source-to-detector were 100 cm and 150 cm. respectively.

B. Low-Dose Projection Simulation

In order to evaluate our self-supervised denoising approach, we generated various low-dose projections. Using Poisson distribution in (2), we simulated 5%, 10%, 20%, and 40% doses of the acquired x-ray projections. With the projections of each simulated low-dose scan, we trained the denoising network under self-supervision as described in the previous section. After training, the same low-dose projection views, which were used during the training process, were fed into the learned network for prediction. Note that the scenario is equivalent to one low-dose CBCT study of a patient being available for both training and prediction.

C. Model Implementation

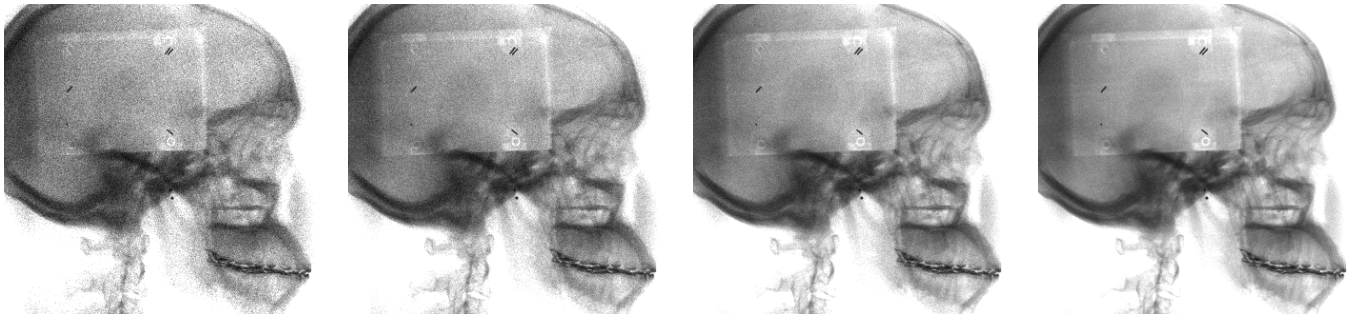
We implemented the denoising network in PyTorch running on NVIDIA TITAN Xp GPUs. The neural networks were initialized with normal distribution and trained by Adam optimizer with 0.5 of momentum during 100 epochs. The learning rate was set to 2×10^{-4} and linearly decayed to zero. The batch size was set to 8 for training and 1 for inference. For training and inference, the network used full-size 1024×768 projections for end-to-end processing.

D. Image Reconstruction

Since the projection views were acquired by a single circular CBCT scan, we used FDK algorithm to produce 3D volumes for both simulated low-dose and denoised projections. The dimension of CBCT volume was 512×512×256 voxels. The FOV of CBCT images was 256 mm with the pixel size of 0.5 mm. The CBCT images were reconstructed in 1.5 mm slice thickness and 1 mm slice interval.

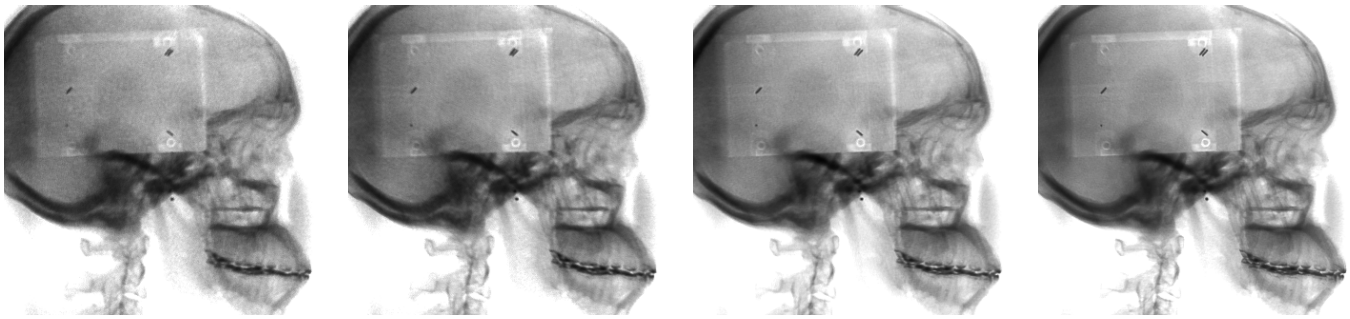
E. Qualitative and Quantitative Analysis

After training, the simulated low-dose projections were again fed into the denoising network for inference. The noisy input projections and the corresponding denoised projections are shown in Figures 1 and 2, respectively. We also show the reconstructed CBCT images with the conventional FDK algorithm in Figures 3 and 4, respectively.



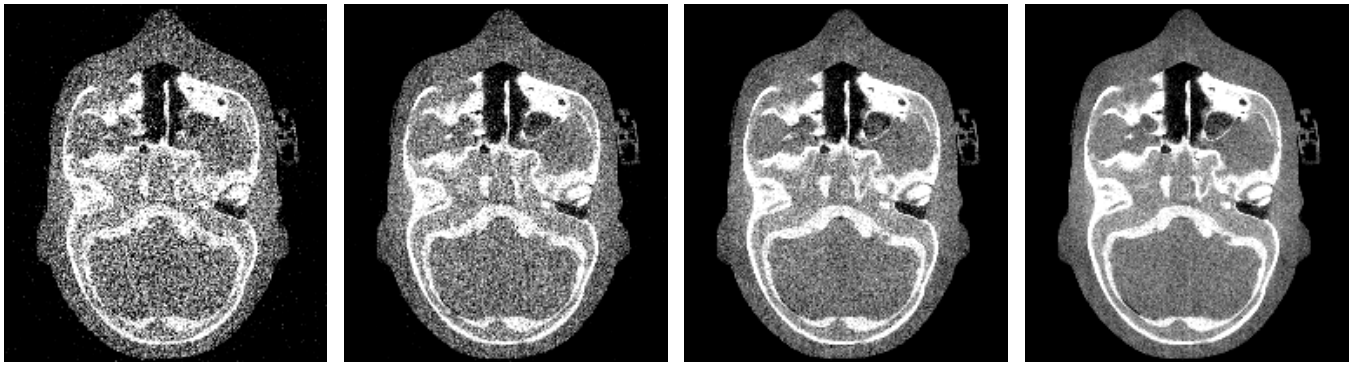
(a) 5% dose (b) 10% dose (c) 20% dose (d) 40% dose

Fig. 1. Simulated low-dose projection views. Each projection view was normalized by the maximum incident photon number. Display window is [0, 0.1]



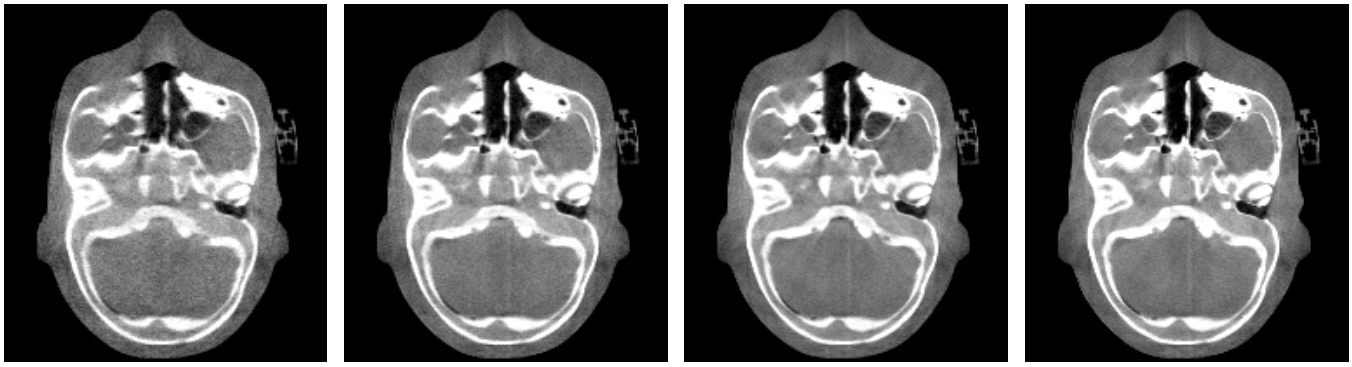
(a) 5% dose (b) 10% dose (c) 20% dose (d) 40% dose

Fig. 2. Denoised low-dose projection views from our self-supervised learning model. Each projection view was normalized by the maximum incident photon number. Display window is [0, 0.1].



(a) 5% dose (b) 10% dose (c) 20% dose (d) 40% dose

Fig. 3. Reconstructed CBCT images from simulated low-dose projections. Display window is [-500, 900] HU.



(a) 5% dose (b) 10% dose (c) 20% dose (d) 40% dose

Fig. 4. Reconstructed CBCT images from denoised low-dose projections. Display window is [-500, 900] HU.

	5% dose	10% dose	20% dose	40% dose
Noisy inputs	27.1±1.4	29.6±1.5	34.1±1.3	37.9±1.3
Denoised results	31.7±3.2	32.3±3.5	35.9±2.9	35.9±3.5

TABLE I

PSNR (MEAN±STD) OF SIMULATED AND DENOISED LOW-DOSE PROJECTIONS.

	5% dose	10% dose	20% dose	40% dose
Noisy inputs	.663±.048	.753±.040	.845±.028	.924±.015
Denoised results	.888±.031	.909±.030	.932±.019	.944±.015

TABLE II

SSIM (MEAN±STD) OF SIMULATED AND DENOISED LOW-DOSE PROJECTIONS.

	5% dose	10% dose	20% dose	40% dose
Noisy inputs	19.5±1.2	22.6±1.3	26.1±1.3	30.4±1.3
Denoised results	28.5±1.3	30.4±1.1	31.1±1.6	32.1±1.4

TABLE III

PSNR (MEAN±STD) OF RECONSTRUCTED IMAGES FROM SIMULATED AND DENOISED LOW-DOSE PROJECTIONS.

	5% dose	10% dose	20% dose	40% dose
Noisy inputs	.699±.055	.782±.028	.835±.022	.899±.015
Denoised results	.829±.026	.863±.019	.874±.021	.904±.014

TABLE IV

SSIM (MEAN±STD) OF RECONSTRUCTED IMAGES FROM SIMULATED AND DENOISED LOW-DOSE PROJECTIONS.

For quantitative comparison between the simulated and denoised low-dose projections, we calculated peak signal-to-noise ratio (PSNR) and structural similarity index measure (SSIM) of the simulated and denoised projections with respect to the acquired high-dose projections as reference. The PSNR and SSIM averaged over 680 views are listed for the simulated low-dose projections and the corresponding denoised projections in Tables I and II, respectively.

For performance comparison in image domain, we also measured PSNR and SSIM of the reconstructed images from the simulated and denoised projection with respect to the high-quality images reconstructed from the acquired high-dose projections. The PSNR and SSIM averaged over 256 slices of each CBCT volume, which were reconstructed from the simulated low-dose and denoised projections, are listed in Tables III and IV, respectively.

IV. CONCLUSION

In this paper, we propose a self-supervised denoising approach which learns noise pattern in projection views for low-dose CBCT imaging. To capture the spatially varying noise pattern of low-dose projections, we extend a contemporary network to cover full-size CBCT projections. Through phantom experiments with simulated low-dose projections, we showed that the proposed self-supervised denoising approach can perform view-wise learning with the low-dose projections alone.

REFERENCES

[1] D. J. Brenner and E. J. Hall, "Computed tomography-an increasing source of radiation exposure," *N. Engl. J. Med.*, vol. 357, pp. 2277–2284, 2007.

[2] D. P. Naidich, C. H. Marshall, C. Gribbin, R. S. Arams, and D. I. McCauley, "Low-dose CT of the lungs: preliminary observations." *Radiology*, vol. 175, no. 3, pp. 729–731, 1990.

[3] Y. Sagara, A. K. Hara, W. Pavlicek, A. C. Silva, R. G. Paden, and Q. Wu, "Abdominal CT: comparison of low-dose CT with adaptive statistical iterative reconstruction and routine-dose CT with filtered back projection in 53 patients," *American Journal of Roentgenology*, vol. 195, no. 3, pp. 713–719, 2010.

[4] D. L. Snyder, M. Miller, L. J. Thomas Jr, D. G. Politte *et al.*, "Noise and edge artifacts in maximum-likelihood reconstructions for emission tomography," *IEEE Trans. Med. Imag.*, vol. 6, no. 3, pp. 228–238, 1987.

[5] J. Hsieh, *Computed Tomography: Principles, Design, Artifacts, and Recent Advances*. SPIE, 2009.

[6] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," in *Advances in Neural Information Processing Systems*, 2012, pp. 1097–1105.

[7] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," in *International Conference on Learning Representations*, 2015.

[8] I. Goodfellow, Y. Bengio, and A. Courville, *Deep learning*. MIT press, 2016.

[9] H. Chen, Y. Zhang, W. Zhang, P. Liao, K. Li, J. Zhou, and G. Wang, "Low-dose CT via convolutional neural network," *Biomedical optics express*, vol. 8, no. 2, pp. 679–694, 2017.

[10] K. Choi, S. W. Kim, and J. S. Lim, "Real-time image reconstruction for low-dose CT using deep convolutional generative adversarial networks (GANs)," in *SPIE Medical Imaging 2018: Physics of Medical Imaging*, vol. 10573, 2018, p. 1057332.

[11] H. Shan, Y. Zhang, Q. Yang, U. Kruger, M. K. Kalra, L. Sun, W. Cong, and G. Wang, "3-D convolutional encoder-decoder network for low-dose CT via transfer learning from a 2-D trained network," *IEEE Trans. Med. Imag.*, vol. 37, no. 6, pp. 1522–1534, 2018.

[12] K. Choi, J. S. Lim, and S. Kim, "StatNet: Statistical Image Restoration for Low-Dose CT using Deep Learning," *IEEE J. Sel. Topics Signal Process.*, vol. 14, no. 6, pp. 1137–1150, 2020.

[13] J. Lehtinen, J. Munkberg, J. Hasselgren, S. Laine, T. Karras, M. Aittala, and T. Aila, "Noise2Noise: Learning Image Restoration without Clean Data," in *International Conference on Machine Learning*, 2018, pp. 2971–2980.

[14] J. Batson and L. Royer, "Noise2Self: Blind Denoising by Self-Supervision," in *International Conference on Machine Learning*, 2019, pp. 524–533.

[15] A. Krull, T.-O. Buchholz, and F. Jug, "Noise2void-learning denoising from single noisy images," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2019, pp. 2129–2137.

[16] S. Laine, T. Karras, J. Lehtinen, and T. Aila, "High-quality self-supervised deep image denoising," in *Advances in Neural Information Processing Systems*, 2019, pp. 6970–6980.

[17] K. Zhang, W. Zuo, Y. Chen, D. Meng, and L. Zhang, "Beyond a gaussian denoiser: Residual learning of deep cnn for image denoising," *IEEE Transactions on Image Processing*, vol. 26, no. 7, pp. 3142–3155, 2017.

[18] J. Johnson, A. Alahi, and L. Fei-Fei, "Perceptual losses for real-time style transfer and super-resolution," in *European conference on computer vision*. Springer, 2016, pp. 694–711.

[19] P. Isola, J.-Y. Zhu, T. Zhou, and A. A. Efros, "Image-to-image translation with conditional adversarial networks," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2017, pp. 1125–1134.

[20] J.-Y. Zhu, T. Park, P. Isola, and A. A. Efros, "Unpaired image-to-image translation using cycle-consistent adversarial networks," in *Proceedings of the IEEE International Conference on Computer Vision*, 2017, pp. 2223–2232.