Sparsier2Sparse: Weakly-supervised learning for streak artifact reduction with unpaired sparse-view CT data

Seongjun Kim, Byeongjoon Kim, and Jongduk Baek

School of Integrated Technology and Yonsei Institute of Convergence Technology, Yonsei University, Incheon

ABSTRACT

Sparse-view computed tomography (CT) becomes a major concern in the medical imaging field due to its reduced X-ray radiation dose. Recently, various convolutional neural network (CNN)-based approaches have been proposed, requiring the pairs of full and sparse-view CT images for network training. However, these paired data acquisition is impractical or difficult in clinical practice. To handle this problem, we propose the weakly-supervised learning for streak artifact reduction with unpaired sparse-view CT data. For CNN training dataset, we generate the pairs of input and target images from the given sparse-view CT data. Then, we iteratively apply the trained network to given sparse-view CT images and acquire the prior images. As the success factor of our novel framework, we estimate the original streak artifacts in the given sparse-view CT images from the prior images and subtract the estimated streak artifacts from the given sparse-view CT images. As a result, the proposed method has the best performance of lesion detection compared to the other methods.

Keywords: Computed tomography, Convolutional neural network, Weakly-supervised learning, Sparse-view CT

1. INTRODUCTION

Computed tomography (CT) has been widely used for disease diagnosis in the medical imaging field due to its detailed observations of anatomical structures. However, the X-ray radiation exposure can increase a potential risk of cancers to the patients. To reduce the X-ray radiation dose, decreasing the number of projection views during CT scan, called sparse-view CT, can be implemented, whereas the severe streak artifacts are generated in the reconstructed images.

To overcome this, several iterative reconstruction (IR) methods, which iterativley optimize the both CT data fidelity and the image regularization terms, have been developed for sparse-view CT. For the prior of image regularization term, total variation[1,2] is often utilized. However, IR calculates the forward and backprojection during optimization, requiring a huge amount of computational cost. Moreover, it is not easy to balance the parameters of fidelity and regularization terms for various imaging tasks. Therefore, the usage of IR methods is limited in medical imaging field, where the real time applications are required.

Recently, deep learning-based approaches have shown promising results in sparse-view CT reconstruction. Inspired by convolutional neural networks (CNNs), the image-domain method[3–5] trains the spatial information of streak artifacts distribution and reduces the streak artifacts effectively, whereas it also often reduces the signals overlapped by streak artifacts together. To overcome this limitation, the hybrid-domain approach,[6–8] which contains both image and sinogram domain, not only learns the image prior but also utilizes the information of measured projection data during network training. From this, the hybrid-domain method improves the performance of streak artifact reduction while preserving the edge sharpness better than when only the image-domain is used.

Although the above CNN-based methods showed good performance, they are fully-supervised approaches that require the paired full and sparse-view CT images, where the anatomical structures are identical. In practice,

Seongjun Kim: E-mail: 2sjkim@gmail.com
Byeongjoon Kim: E-mail: bjkim2006@naver.com
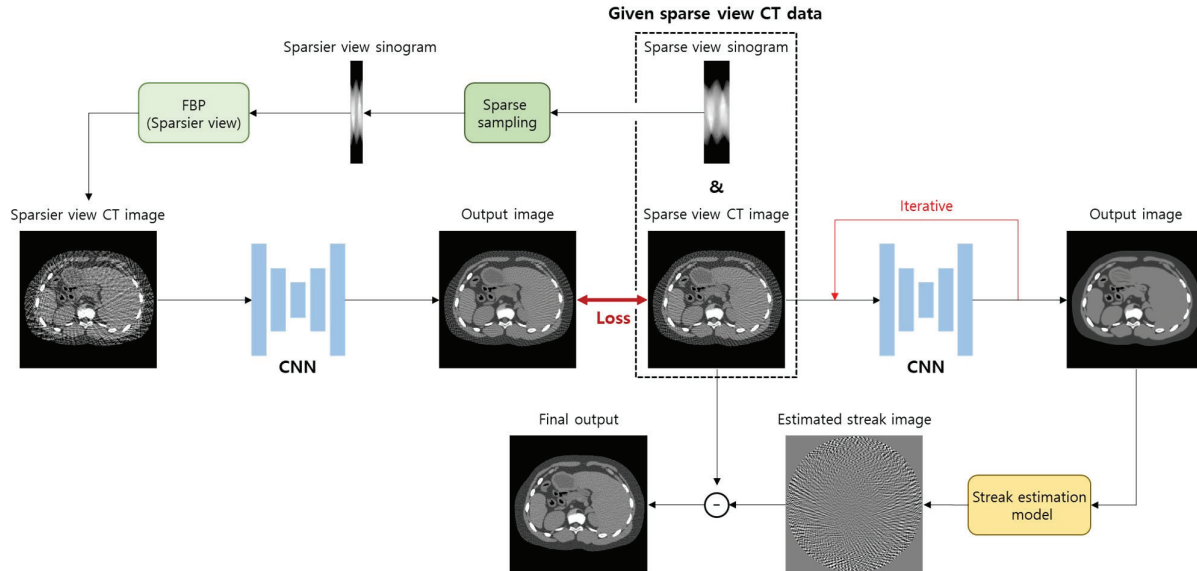Jongduk Baek: jongdukbaek@yonsei.ac.kr

Figure 1. The overall schematic of the proposed method. FP: forward projection; FBP: filtered back projection

it is not feasible to acquire the dataset of those pairs from the patients due to the principle of ALARA (as low as reasonably achievable). In this reason, to acquire the full and sparse-view CT image pairs, most of methods generate the sparse-view CT images by a computer simulation using the full view CT data. However, if only the sparse-view CT data are given, it is impossible to produce full view CT images that are anatomically identical. This is because image reconstruction in sparse-view CT is an undetermined inverse problem that has no general solution.[9]

To tackle this problem, we propose a weakly-supervised learning for streak artifact reduction when the only unpaired sparse-view CT data are given. We acquire the pairs of data for CNN training from the given sparse-view CT data and train the network. We then apply trained network to the given sparse-view CT images iteratively. For the success of our framework, we perform streak estimation model for signal preservation.

2. METHODS

Figure 1 shows the schematic diagram of the proposed method for reducing streak artifacts with unpaired sparse-view CT data.

At times it may be desired, for formatting reasons, to break a line without starting a new paragraph. This situation may occur, for example, when formatting the article title, author information, or section headings. Line breaks are inserted in LaTeX by entering \\ or \linebreak in the LaTeX source file at the desired location.

2.1 Weakly-supervised learning

For CNN training to reduce streak artifacts, it is required to set the images with strong and weak streak artifacts as input and target, respectively. Since the only unpaired sparse-view CT data are given in our scenario, we should acquire pairs of the input and target from given sparse-view CT data. For this, we generate an image that has stronger streak artifacts than the original streak artifacts in given sparse-view CT image, which is denoted as a sparsier-view CT image. We then regard the sparsier and sparse-view CT images as the input and target for CNN training, respectively.

Here, the method of sparsier-view CT image generation is described in the following procedures. To make streak artifacts stronger, it can be achieved by decreasing the number of projection views because the strength of the streak artifacts is dependent on the number of projection views. Therefore, we down-sampled the given sparse-view sinogram by half and acquire a sparsier-view sinogram. Then, we reconstruct the sparsier-view CT image from the sparsier-view sinogram with the filtered backprojection (FBP) algorithm.[10] Note that we set

down-sampling ratio as 2 to ease CNN training because the spasier view CT image will be significantly corrupted when the down-sampling ratio is bigger than 2.

Let $\hat{x}$ and x be the sparsier-view CT image and sparse-view CT image, respectively. Since the projection data of $\hat{x}$ is extracted and down-sampled from that of x, the streak artifacts in $\hat{x}$ will be amplified while keeping the directionality of the original streak artifacts in x. Therefore, we set a pair of $\hat{x}$ and x as CNN training dataset. We then train CNN with the mean squared error (MSE). The MSE is defined as

$$L_{MSE} = \mathbb{E}\big[\|f(\hat{x};\theta) - x\|^2\big] \tag{1}$$

where $f(\cdot;\theta)$ and θ denote the network operator and network parameters, respectively.

2.2 Iterative streak artifact reduction

Since the network trains to reduce streak artifacts by predicting the original streak artifacts in x from the amplified streak artifacts in $\hat{x}$, we iteratively apply the trained network to a given x to reduce the streak artifacts. In the first iteration, we set x as the network input and acquire the network output $f(x;\theta)$. Since a single iteration is not sufficient to reduce streak artifacts in x, more than two iterations are conducted. Note that the output of the previous iteration was set as the input of the current iteration. We perform this processing until the streak artifacts in x are reduced to the level we want. Note that the number of iterations can be flexibly controlled and it was set to a maximum of five in this work. We will notate y as the resulting image of the last iteration in this step.

2.3 Streak artifact estimation

Applying the trained network operator several times works to reduce streak artifacts rather than preserve the anatomical structure. Therefore, the streak artifacts are significantly reduced in y, but over-smoothing and blurring can occur, resulting in low performance of signal detection. In this reason, it is inadequate to utilize y itself as the final result. To preserve the fine details that may have been blurred, we estimate the original streak artifacts in x from y.

For streak estimation model, we first reconstruct full and sparse-view images from y by sequentially applying forward projection and FBP algorithm with full and sparse-view projection data, respectively. These full view image and sparse-view image are denoted as y_{full} and y_{sparse}, respectively. The estimated streak image is generated by subtracting y_{full} from y_{sparse}. We then acquire the final output by subtracting the estimated streak image from x.

2.4 Training configurations

We used the U-net structure[11] because it is well-known for effectively extracting the widely distributed streak artifacts in the reconstructed images. The network was optimized by Adam optimizer[12] with default parameters and the learning rate 1e-4. To increase the datasets for CNN training, we implemented data augmentation that randomly flips and rotates the cropped images, where the images of 256 by 256 pixels are cropped from the original images of 512 by 512 pixels. We trained the network until 100 epochs and set batch size as 4.

3. DATASETS AND EXPERIMENTS

3.1 XCAT dataset

We used abdomen and thorax regions of XCAT simulation dataset in our work. A total of 9690 slices was extracted from the 28 XCAT phantoms. We generated XCAT images in fan-beam geometry system and the simulation parameters are summarized in Table 1. During image generation, the forward projection was implemented using Siddon's ray-driven algorithm[13] and the FBP algorithm was conducted with a Ram-Lak filter. For CNN dataset, we used 25 phantoms (8720 slices) and 3 phantoms (970 slices) for training and test set, respectively. The different details between sparse and full view CT images are explained in the following section.

Table 1. Fan geometry simulation parameters

Parameters	Values
Source to iso-center distance	$50cm$
Source to detector distance	$50cm$
Detector pixel size	512 x 1
Detector cell size	$0.2cm^2$ x $0.1cm^2$
Reconstructed matrix size	512 x 512
Reconstructed pixel size	$0.0628cm^2$ x $0.0628cm^2$

3.2 Data generation

We generated sparse and full view CT images with 128 and 512 projection views, respectively, and each projection angle was equally spaced over 360 degrees. For Poisson noise, we set the number of incident X-ray photons as 10^6 for full view CT images. However, since the number of projection views in the sparse-view CT images is a quarter of full view CT images, the noise level of sparse-view CT images will be four times higher than that of full view CT images when the number of incident X-ray photons are the same. To make the noise level equal, we set the number of incident X-ray photons as 4×10^6 for sparse-view CT images. Note that the only unpaired sparse-view CT images are given for CNN training and the full view CT images are only used in testing phase.

3.3 Comparison methods

We additionally conducted a simple linear interpolation method and a fully-supervised learning[11] for comparison. The linear interpolation method estimated the missing view data of sparse-view sinogram by applying linear interpolation in view direction. For fully-supervised learning, we assumed that the paired of sparse and full view CT images were given for input and target of CNN training, respectively. The training configurations were the same as that we used in the proposed method.

4. RESULTS

Figure 2 shows the resulting images on XCAT dataset. The linear interpolation image showed less streak artifacts than 128 view FBP image through the estimation of missing view data, whereas the secondary artifacts such as edge distortion were produced due to the interpolation errors. The fully-supervised method reduced the streak artifacts and noise significantly. However, it can be seen that the lesions indicated by red and yellow arrows in ROIs of the fully-supervised method were blurred, leading to poor performance of visibility of those signals. In contrast, the proposed method produced the best visual similarity of ROIs to those of 512 view FBP image while preserving the edge sharpness. For the proposed method with iteration 1, its difference image had more edge errors than the results of other iterations. However, these errors disappear as more iterations are conducted.

Table 2 summarizes the average and standard deviation of normalized root MSE (NRMSE) and structural similarity index (SSIM)[14] on XCAT testset. We observed that the fully-supervised method showed the best scores of NRMSE and SSIM. For the proposed method, iteration 2 had the lowest NRMSE and highest SSIM scores. Although the proposed method with iteration 2 showed slightly worse scores than the fully-supervised method, visual inspection confirms that the proposed method produces better image texture without sacrificing the detectability of lesions.

To examine the effect of the streak estimation model in the proposed method, we did not apply the streak estimation model after the iterative streak artifact reduction procedure and the results are shown in figure 3. As the number of iterations increases, although the streak artifacts are gradually reduced, it can be seen that the images were getting blurred. For the ROIs of more than iteration 2, the signals were lost and the edge was also over smoothed. Although the results of iteration 1 preserved the signal indicated by red arrow, the streak artifacts were not reduced enough. From this, we chose to use iteration 2 to effectively reduce streak artifacts while preserving edge and signal shapes.

128 view   Linear interpolation   Fully-supervised   Proposed; Iteration 1   Proposed; Iteration 2   Proposed; Iteration 3   Proposed; Iteration 4   Proposed; Iteration 5   512 view
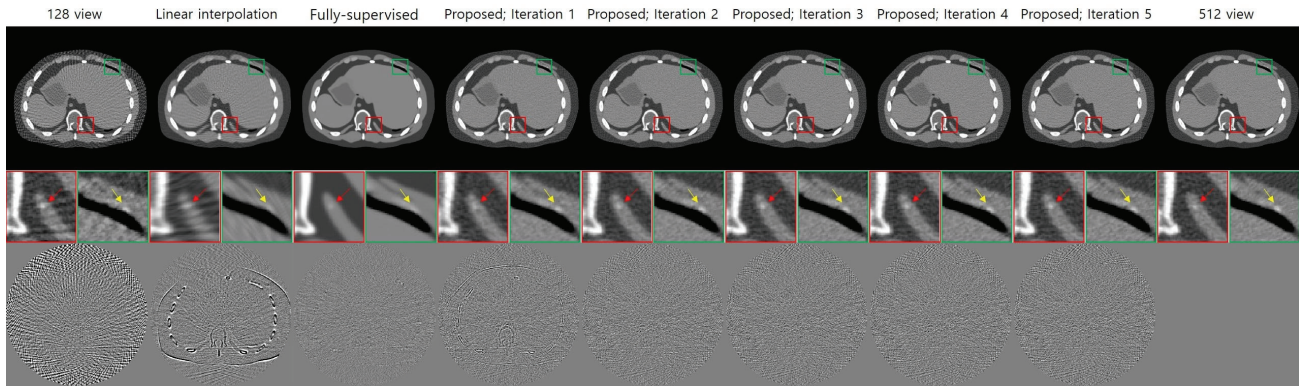
Figure 2. The reconstructed images with XCAT dataset. The display window of field of view (FOV) and zoomed region of interest (ROI) images is (-240,320) in HU. The display window of difference images is (-100,100) in HU.

128 view   Iteration 1   Iteration 2   Iteration 3   Iteration 4   Iteration 5   512 view
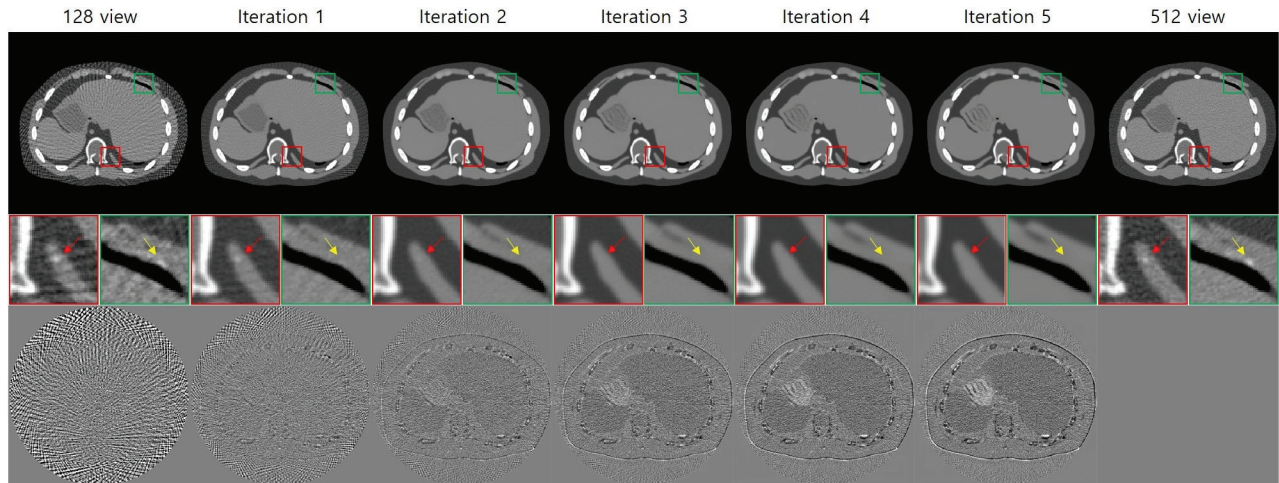
Figure 3. Prior images after conducting different number of iterations in the proposed method. The display window of FOV and ROI images is (-240,320) in HU. The display window of difference images is (-100,100) in HU.

Table 2. The quantitative evaluations of XCAT dataset

| Metrics | 128 view | Linear interpolation | Fully-supervised | Proposed | | | | |
				Iteration 1	Iteration 2	Iteration 3	Iteration 4	Iteration 5
NRMSE	0.0268 ± 0.0042	0.0157 ± 0.0036	$\mathbf{0.0064 \pm 0.0013}$	0.0122 ± 0.0015	$\mathbf{0.0102 \pm 0.0012}$	0.0104 ± 0.0012	0.0108 ± 0.0012	0.0111 ± 0.0012
SSIM	0.6838 ± 0.0484	0.8988 ± 0.0340	$\mathbf{0.9659 \pm 0.0133}$	0.9004 ± 0.0135	$\mathbf{0.9101 \pm 0.0115}$	0.9019 ± 0.0117	0.8938 ± 0.0119	0.8870 ± 0.0121

5. CONCLUSION

In this work, we proposed the weakly-supervised learning for streak artifact reduction with unpaired sparse-view CT data. From the results, the proposed method achieved the best performance in preserving lesions while reducing streak artifacts compared to the other results. By overcoming the difficulty of acquiring the pairs of sparse and full view CT images in practice, we expect that our novel framework can be utilized successfully in medical imaging field.

ACKNOWLEDGMENTS

REFERENCES

[1] Park, J. C., Song, B., Kim, J. S., Park, S. H., Kim, H. K., Liu, Z., Suh, T. S., and Song, W. Y., "Fast compressed sensing-based cbct reconstruction using barzilai-borwein formulation for application to on-line igrt," *Medical physics* **39**(3), 1207–1217 (2012).

[2] Sidky, E. Y. and Pan, X., "Image reconstruction in circular cone-beam computed tomography by constrained, total-variation minimization," *Physics in Medicine & Biology* **53**(17), 4777 (2008).

[3] 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 transactions on medical imaging* **37**(6), 1407–1417 (2018).

[4] Han, Y. and Ye, J. C., "Framing u-net via deep convolutional framelets: Application to sparse-view ct," *IEEE transactions on medical imaging* **37**(6), 1418–1429 (2018).

[5] Jin, K. H., McCann, M. T., Froustey, E., and Unser, M., "Deep convolutional neural network for inverse problems in imaging," *IEEE Transactions on Image Processing* **26**(9), 4509–4522 (2017).

[6] Liang, K., Yang, H., and Xing, Y., "Comparison of projection domain, image domain, and comprehensive deep learning for sparse-view x-ray ct image reconstruction," *arXiv preprint arXiv:1804.04289* (2018).

[7] Zheng, A., Gao, H., Zhang, L., and Xing, Y., "A dual-domain deep learning-based reconstruction method for fully 3d sparse data helical ct," *Physics in Medicine & Biology* **65**(24), 245030 (2020).

[8] Lee, D., Choi, S., and Kim, H.-J., "High quality imaging from sparsely sampled computed tomography data with deep learning and wavelet transform in various domains," *Medical physics* **46**(1), 104–115 (2019).

[9] Wu, W., Hu, D., Niu, C., Yu, H., Vardhanabhuti, V., and Wang, G., "Drone: Dual-domain residual-based optimization network for sparse-view ct reconstruction," *IEEE Transactions on Medical Imaging* (2021).

[10] Hsieh, J., [*Computed tomography: principles, design, artifacts, and recent advances*], vol. 114, SPIE press (2003).

[11] Kim, B., Han, M., Shim, H., and Baek, J., "A performance comparison of convolutional neural network-based image denoising methods: The effect of loss functions on low-dose ct images," *Medical physics* **46**(9), 3906–3923 (2019).

[12] Kingma, D. P. and Ba, J., "Adam: A method for stochastic optimization," *arXiv preprint arXiv:1412.6980* (2014).

[13] Siddon, R. L., "Fast calculation of the exact radiological path for a three-dimensional ct array," *Medical physics* **12**(2), 252–255 (1985).

[14] 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**(4), 600–612 (2004).