RESEARCH ARTICLE

Sparsier2Sparse: Self-supervised convolutional neural network-based streak artifacts reduction in sparse-view CT images

Seongjun Kim[1] | **Byeongjoon Kim[2]** | **Jooho Lee[2]** | **Jongduk Baek[2,3]**

[1]School of Integrated Technology, Yonsei University, Incheon, South Korea

[2]Department of Artificial Intelligence, College of Computing, Yonsei University, Seoul, South Korea

[3]Bareunex Imaging, Inc., Seoul, South Korea

Correspondence
Jongduk Baek, Department of Artificial Intelligence, College of Computing, Yonsei University, Seoul, South Korea.
Email: jongdukbaek@yonsei.ac.kr

Abstract

Background: Sparse-view computed tomography (CT) has attracted a lot of attention for reducing both scanning time and radiation dose. However, sparsely-sampled projection data generate severe streak artifacts in the reconstructed images. In recent decades, many sparse-view CT reconstruction techniques based on fully-supervised learning have been proposed and have shown promising results. However, it is not feasible to acquire pairs of full-view and sparse-view CT images in real clinical practice.

Purpose: In this study, we propose a novel self-supervised convolutional neural network (CNN) method to reduce streak artifacts in sparse-view CT images.

Methods: We generate the training dataset using only sparse-view CT data and train CNN based on self-supervised learning. Since the streak artifacts can be estimated using prior images under the same CT geometry system, we acquire prior images by iteratively applying the trained network to given sparse-view CT images. We then subtract the estimated steak artifacts from given sparse-view CT images to produce the final results.

Results: We validated the imaging performance of the proposed method using extended cardiac-torso (XCAT) and the 2016 AAPM Low-Dose CT Grand Challenge dataset from Mayo Clinic. From the results of visual inspection and modulation transfer function (MTF), the proposed method preserved the anatomical structures effectively and showed higher image resolution compared to the various streak artifacts reduction methods for all projection views.

Conclusions: We propose a new framework for streak artifacts reduction when only the sparse-view CT data are given. Although we do not use any information of full-view CT data for CNN training, the proposed method achieved the highest performance in preserving fine details. By overcoming the limitation of dataset requirements on fully-supervised-based methods, we expect that our framework can be utilized in the medical imaging field.

KEYWORDS
computed tomography, convolutional neural network, self-supervised learning, sparse-view CT

1 | INTRODUCTION

Computed tomography (CT) is a widely utilized imaging system for disease diagnosis with detailed visualization of fine structures. It reconstructs interior organs of diverse regions by detecting X-ray photons pene-

trated from the body and utilizing acquired projection data. However, radiation exposure to the patients during scanning may lead to the breaking of molecular bonds,[1] potentially increasing the risk of cancers. To handle this problem, various CT reconstruction techniques for reducing radiation dose such as low-dose CT,[2] limited-angle CT,[3] and sparse-view CT[4–9] have been developed in the medical imaging field. Among these low-dose

Seongjun Kim is the first author.

CT imaging protocols, we focus on sparse-view CT that lowers the number of projection views for CT scanning trajectory, achieving both reduced scanning time and radiation dose. However, the reconstructed images of sparse-view CT suffer from severe streak artifacts caused by under-sampled projection data,[4,10] and it is challenging to restore full-view CT images from sparse-view CT images because sparse-view CT reconstruction is an undetermined inverse problem with no general solution.[8]

As the traditional method to reduce streak artifacts, iterative reconstruction (IR) has been developed. IR approach iteratively optimizes both data fidelity and image regularization terms, where total variation (TV)[11,12] norm is commonly used as image prior to preserve edge representation[8,13–15] while reducing streak artifacts. However, due to multiple calculations of forward and backprojection during optimization, IR requires a significant computational cost, and thus consumes a long computation time. It is also not a simple task to adjust weighting parameters on data fidelity and regularization terms for different CT imaging schemes and body parts.[4] As a result, IR approach is not proper to be used in real-time clinical applications.

Recently, various convolutional neural network (CNN) methods based on deep learning (DL) have been addressed for sparse-view CT reconstruction, overcoming the limitations of the existing IR methods. The image-domain approach[5–7] captures the spatial correlation of globally distributed streak artifacts and reduces them, but fine details in anatomical structures can also be reduced together when they are overlapped by severe streak artifacts. To better preserve the original image details, the dual-domain CNN method, which utilizes both image and sinogram domains, has been proposed.[16–18] This approach realizes high-quality images by learning the patterns of streak artifacts and exploiting the information in measured data during the training phase. Several methods that combine CNN and IR have also been proposed. Chen et al.[4] developed learned experts' assessment-based reconstruction network (LEARN) for sparse-view CT, where IR framework was unfolded into multiple steps with CNNs. LEARN utilized DL-based prior as the image regularization term and simultaneously optimized data fidelity and regularization terms to enhance generalization performance. AirNet[19] was also designed similarly to LEARN, but it used filtered backprojection (FBP) operator in place of simple backprojection and utilized dense connectivity between iterations, further improving the image quality and being robust to the dataset. Wu et al.[8] trained CNN in image and sinogram domains and used the awareness module with IR framework utilizing the output of CNN for data fidelity and regularization terms to decrease the discrepancy between the learned prior and measured data. Wu et al.[20] integrated deep embedding-attention-refinement modules with the residual error feedback

mechanism, which were applied both on sinogram and image domains. Zhang et al.[9] trained the first image-domain CNN and utilized its output as a prior image for regularization term in IR framework. After that, the second image-domain CNN is trained to reduce the streak artifacts. On the other hand, the method including both CNN and transformer has also been suggested. Pan et al.[21] designed MIST-net to reduce streak artifacts by training not only a CNN but also a transformer-based network to capture long-range features, as CNNs concentrate local image information due to its limited receptive field. Although these aforementioned methods showed promising results, they were based on fully-supervised learning that requires pairs of sparse-view and full-view CT images with identical anatomical structures for network training.

However, for fully-supervised learning, scanning the patients twice under different projection views (i.e., full views and sparse views) should be conducted, which is not feasible in real practice because it is against the as low as reasonably achievable (ALARA) principle. Moreover, the position of the patients would be different during the two scans because of movement, thereby producing low-quality training data. Alternately, we can generate sparse-view CT images by using computer simulation of CT geometry when full-view CT images are given. However, high-quality full-view CT images may not be always available, especially for cardiac or pediatric imaging. In cardiac imaging, reconstructed images are often blurred by rapid heart rate.[22] Thus, reducing scanning time with sparse-view CT is an effective way to address this problem.[23,24] In pediatric imaging, the total radiation dose is reduced using low-dose or sparse-view CT because radiation exposure can be more harmful to pediatric patients than adults.[25]

Therefore, further investigation is required on training CNN using only sparse-view CT images based on self-supervised learning. In the image-denoising field, several self-supervised learning approaches have been proposed. Moran et al.[26] proposed Noisier2Noise, which synthesized additional noise on given original noisy images to generate noisier images, and then trained a neural network from noisier images to noisy images. Building on Noisier2Noise, Kim et al.[15] employed a physics-based CT noise generation module to synthesize lower-dose CT images for CNN training and proposed a progressive denoising framework to further improve the image-denoising performance.

Inspired by these self-supervised denoising methods,[15,26] we aim to reduce streak artifacts in sparse-view CT images with the same purpose of learning CNN without reference images and improving image quality. However, the underlying assumption of Noisier2Noise method is that the noise is statistically random. The previous studies[27,28] showed that the streak artifacts can be modeled by the Gumbel distribution, revealing the statistical characteristics of the
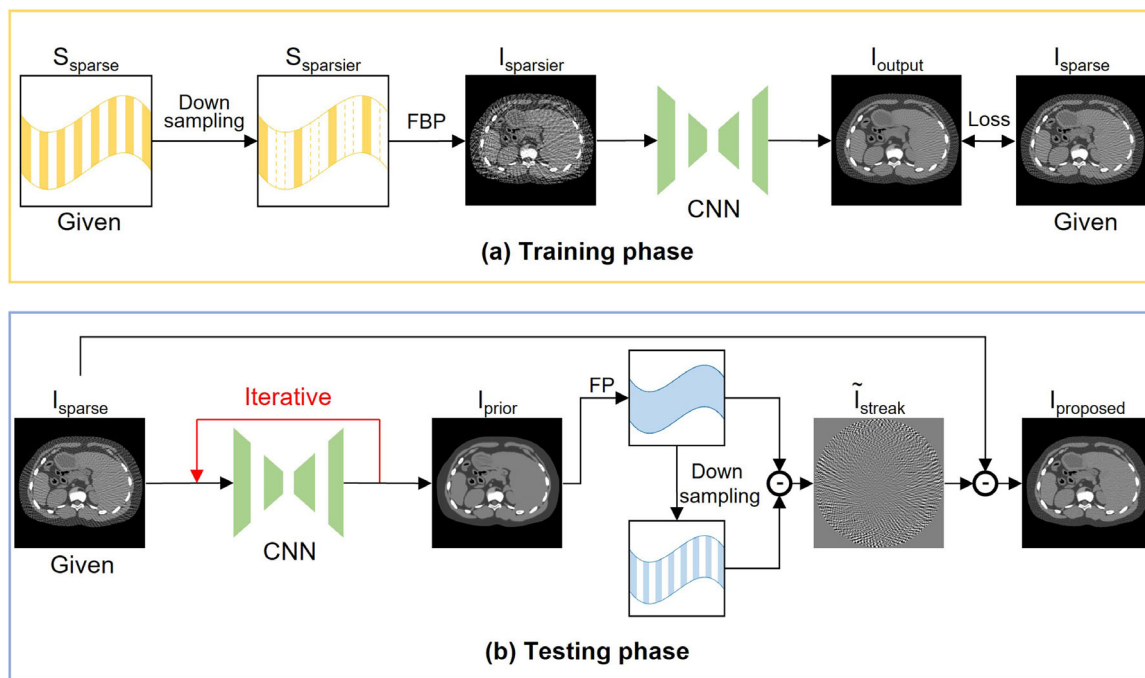
FIGURE 1 The overall diagram of the proposed self-supervised learning-based method for streak artifacts reduction in sparse-view CT. CT, computed tomography.

streak artifacts. Therefore, we treat the streak artifacts as statistical noise and propose a novel framework for streak artifacts reduction by adopting the concept of Noisier2Noise. In the proposed method, we generate the training dataset from given sparse-view CT data and train CNN. The trained CNN processes sparse-view CT images to produce prior image. Then, we estimate the original streak artifacts using prior image and subtract them from given sparse-view CT images. As the extension of our study,[29] we evaluate the performance of the proposed method on both simulation and clinical dataset and compare it with those of various streak artifacts reduction methods using modulation transfer function (MTF) as well as visual comparison.

2 | METHODS

We propose a novel framework to reduce streak artifacts using only sparse-view CT data. Figure 1 illustrates the overall schematic of the proposed method for streak artifacts reduction in sparse-view CT. Our approach consists of two phases: (a) the training phase, and (b) the testing phase. In the training phase, we train a CNN based on a self-supervised approach with the given sparse-view CT data. In the testing phase, we generate a smooth prior image using the trained CNN and estimate the original streak artifacts based on the prior image. We then subtract the estimated streak artifacts from the given sparse-view CT images. The detailed procedures are described in the following steps.

2.1 | Self-supervised learning

Our framework uses only sparse-view CT images (i.e., I_{sparse}) and corresponding sparse-view sinograms (i.e., S_{sparse}) for training CNN. To generate training signals, we down-sample S_{sparse} to acquire sparsier-view sinograms (i.e., $S_{sparsier}$) and then reconstruct sparsier-view CT images (i.e., $I_{sparsier}$). After that, we give $I_{sparsier}$ and I_{sparse} to the network as input and target, respectively. During training, the network learns to remove streak artifacts in input images.

Since our purpose is to reduce the original streak artifacts in I_{sparse} by utilizing sparsier-view CT data (i.e., sparsier than sparse-view CT data), the down-sampling ratio larger than 1 (e.g., 1.2 or 1.5) can be selected to generate $I_{sparsier}$, but we set it as 2 for ease of implementation in this work. Although much larger downsampling ratios (e.g., 4 or 8) can be used, these scales may degrade the quality of the reconstructed image due to the severity of the streak artifacts, and thus we avoid using those scales. Note that $I_{sparsier}$ is reconstructed from $S_{sparsier}$ using the FBP algorithm[1] with a Ram-Lak filter.

The examples of I_{sparse} and $I_{sparsier}$ are shown in Figure 2. From this, the streak artifacts in $I_{sparsier}$ are more severe compared to those of I_{sparse}, while
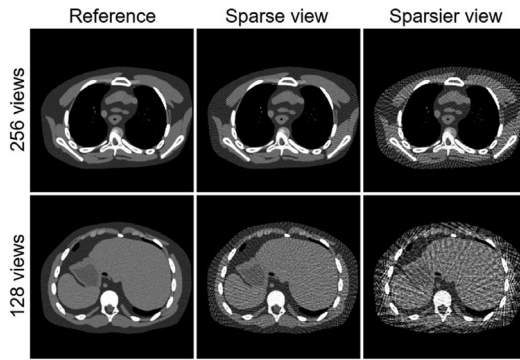
FIGURE 2 Examples of sparse-view and its corresponding sparsier-view CT images on XCAT dataset under different projection views. The display window is [-240 320] in HU. CT, computed tomography; XCAT, extended cardiac-torso.

preserving the directionality of striped patterns of streak artifacts, because the projection data of $S_{sparsier}$ are extracted from those of S_{sparse} during downsampling. We train the image-domain CNN to reduce streak artifacts by considering $I_{sparsier}$ and I_{sparse} as the input and target of the network, respectively. To better lead the network to consider the streak artifacts as the statistical noise during CNN training, we apply various data augmentation techniques (e.g., flip or rotate) to training dataset.

2.2 | Iterative prior generation

Let N_{sparse} and $N_{sparsier}$ be the number of sparse and sparsier projection views, respectively. Due to the inverse relationship between the strength of the streak artifacts and the number of projection views, we define the ratio r as the strength of the streak artifacts between $I_{sparsier}$ and I_{sparse}:

$$r = \frac{N_{sparsier}}{N_{sparse}} \qquad (1)$$

Therefore, the neural network trained by our self-supervised learning method reduces the streak artifacts by a factor of r. Since we aim to obtain full-view CT image (i.e., I_{full}), we reduce the streak artifacts in I_{sparse} by iteratively applying the trained CNN until the strength of reduced streak artifacts reaches that of I_{full}. The required number of iterations (i.e., n) is calculated as:

$$n = \log_r(\frac{N_{sparse}}{N_{full}}) \qquad (2)$$

where N_{full} is the number of full projection views. For example, when N_{full}, N_{sparse}, and $N_{sparsier}$ are given as 512, 128, and 64, respectively, n is calculated as 2. In the first iteration, I_{sparse} is passed through the trained

CNN. From the second iteration, network's output in the previous iteration is set as the input of the network. The output image of the last iteration is denoted as I_{prior}.

2.3 | Estimation of the original streak artifacts

Since both $I_{sparsier}$ and I_{sparse} used for CNN training have already been corrupted by streak artifacts and precise anatomical structures are not visible in those training images, the trained network may have limited performance in preserving fine details. Inspired by the process of streak artifacts regeneration and subtraction suggested by Kim et al.,[30] we used I_{prior} as a prior image to estimate the original streak artifacts in I_{sparse}. The reason why it is possible to estimate streak artifacts is that the same streak artifacts can be regenerated under the same system geometry. Due to the linearity of the FBP algorithm,[31] the original streak artifacts in I_{sparse} can be determined by subtracting I_{full} from I_{sparse}. However, since we do not know any information of I_{full} in our scenario, we generate the pairs of reconstructed full-view and sparse-view CT images from I_{prior}. For this, we first conduct the forward projection of I_{prior} in full-view geometry to obtain full-view projection data, and then down-sample it to generate its corresponding sparse-view projection data. After that, the full-view and sparse-view CT images are reconstructed from full-view and sparse-view projection data using the FBP algorithm, respectively. We then subtract those reconstructed images from each other to acquire the estimated streak artifacts, $\tilde{I}_{streak}$, which can be derived as:

$$\tilde{I}_{streak} = F_{sparse}D(A_{full}I_{prior}) - F_{full}A_{full}I_{prior} \qquad (3)$$

where A_{full} is the matrix for the image modeling system in full-view geometry and D is the downsampling operator that extracts sparse-view projection data from full-view projection data. F_{sparse} and F_{full} represent the matrixes of the FBP operator performed in sparse-view and full-view geometry, respectively. By subtracting $\tilde{I}_{streak}$ from given I_{sparse}, we acquire the final result of the proposed method.

3 | DATASETS AND EXPERIMENTS

3.1 | XCAT and clinical dataset generation

To evaluate the performance of the proposed method, we used the simulation dataset using extended cardiac-torso (XCAT) phantoms[32] and the real clinical dataset from the 2016 AAPM Low-Dose CT Grand Challenge,[33] provided by Mayo Clinic. We generated the projection

data in a fan-beam geometry system using Siddon's ray-driven algorithm.[34] For the XCAT dataset, we added the Poisson noise to the noiseless projection data using 10^6 incident X-ray photons for each detector cell. However, in the case of Mayo Clinic dataset, its original images were already contaminated by the realistic noise, thus we did not add the noise additionally. After the noisy projection data were acquired, we reconstructed full-view and sparse-view CT images from those projection data using the FBP algorithm. Note that the full-view CT images were used only for the qualitative and quantitative evaluation tasks.

For the XCAT dataset, we extracted a total of 4370 slices containing the abdomen and the thorax regions from 13 patients. Among these, 2800 slices from 8 patients were used for training, 600 slices from 2 patients were used for validation, and 970 slices from 3 patients were used for testing. In the clinical dataset, a total of 3000 slices of abdomen parts selected from 10 patients with 300 slices each were divided into 2400 slices from 8 patients, 300 slices from 1 patient, and 300 slices from 1 patient for training, validation, and testing sets, respectively. Note that we separated the training, validation, and testing sets by patients to avoid the potential risk of data contamination.

The detailed descriptions of CT geometry system for image generation are explained as follows. The distances of source to iso-center and source to detector were 500 and 1000 mm, respectively. The number of detector cells was 512 with a size of 2 mm for each cell. The number of projection views was 256, 128, and 64 for sparse-view CT and 512 for full-view CT, respectively, and the rotation angle for each projection view was equal over 360 degrees. The matrix size of the reconstructed images was 512 by 512 pixels with a pixel size of 0.628 by 0.628 mm.

3.2 | Implementation details

For CNN training, we utilized a U-Net structure[30] that has a large receptive field due to its several max pooling layers, which can effectively capture the features of globally distributed streak artifacts.[5,35] However, since the transposed convolution layers in the original U-Net[7] can introduce checkerboard artifacts[36] in the reconstructed images, bilinear upsampling layers were utilized instead of the transposed convolution layers. The network was trained to reduce the mean squared error (MSE) loss between the output of the network and the target image.

Adam optimizer[37] was used to optimize our network with a learning rate of 10^{-4}, using default β_1 and β_2 values of 0.9 and 0.999, respectively. The number of epochs was 100. To address the network overfitting issue, we applied data augmentation during network training, which performs random rotation and flip of the input images. We did not crop the input images and set the batch size as one for fair comparison with AirNet,[19] which will be explained in the next section. The specifications of our computer system's hardware devices for the experiments were composed of Intel Core i7-8700 CPU, 64 GB RAM, and NVIDIA GeForce RTX 4080 with 16 GB memory.

3.3 | Image quality evaluation

We used normalized root MSE (NRMSE) and the structural similarity index (SSIM)[38] for quantitative evaluation. NRMSE calculates the pixel-wise errors and is defined as:

$$NRMSE(\hat{x}, x) = \frac{\|\hat{x} - x\|_2}{max(x) - min(x)} \quad (4)$$

where $\hat{x}$ and x denote the reconstructed image and reference image, respectively. SSIM compares image visual information of luminance, contrast, and structure, which is defined as follows:

$$SSIM(\hat{x}, x) = \frac{2\mu_{\hat{x}}\mu_x + c_1}{\mu_{\hat{x}}^2 + \mu_x^2 + c_1} \cdot \frac{2\sigma_{\hat{x}x} + c_2}{\sigma_{\hat{x}}^2 + \sigma_x^2 + c_2} \quad (5)$$

where $\mu_{\hat{x}}$ and $\sigma_{\hat{x}}$ are the mean value and the variance of $\hat{x}$, respectively, and similar terms corresponding to x are defined in the same way. $\sigma_{\hat{x}x}$ is the covariance of $\hat{x}$ and x. c_1 and c_2 are set to $(0.01L)^2$ and $(0.03L)^2$, where L is a dynamic range which is the largest pixel value of x.

We also examined MTF to evaluate the resolution property of various comparison methods on XCAT dataset for each projection view. MTF was calculated on the regions of interest (ROI) that contained the inserted circular signal with a size of 15 mm diameter and 60 HU contrast, as shown in Figure 3. While following the overall MTF calculation steps proposed in the previous study,[39] we performed Gaussian fitting to a line spread function using a trust-region-reflective least squares algorithm to smooth the curve.

3.4 | Comparison methods

To validate the effectiveness of the proposed method, we compared its imaging performance with that of TV-IR,[11] FBPConveNet,[2] and AirNet.[19] We implemented TV-IR method based on the Gradient-Projection-Barzilai-Borwein algorithm. We set different weights of regularization term for each image to achieve the lowest MSE scores between TV-IR results and full-view CT images. For CNN-based methods, we compared the fully-supervised learning-based FBPConvNet and AirNet. We used sparse-view CT images (for FBPConvNet)
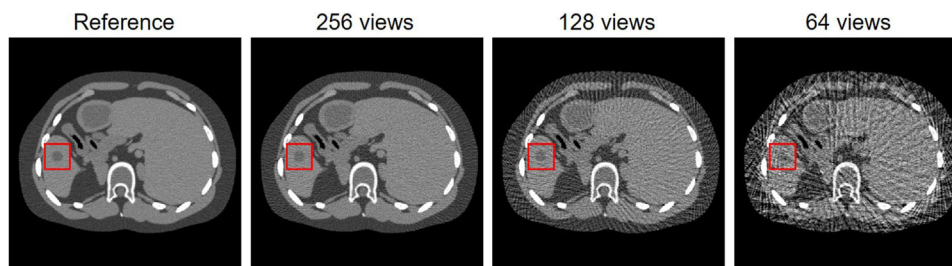
FIGURE 3 Signal-inserted reference and sparse-view CT images with different projection views for MTF calculation. The display window is [-240 320] in HU. CT, computed tomography; MTF, modulation transfer function.

and sparse-view sinograms (for AirNet) as the input and corresponding full-view CT images as the target of these two networks. Note that those sparse-view CT data were identical to the target images in our framework. The other training configurations of FBP-ConvNet and AirNet were the same as those used in the proposed method.

We also compared the proposed method with the recent streak artifact reduction algorithm based on self-supervised neural representation (SAR-SSNR).[30] The difference between the proposed method and SAR-SSNR is that the proposed method produces prior images by training CNN, whereas SAR-SSNR utilizes coordinate-based multilayer perceptron (MLP). We followed the original setup in Kim et al.[30] for implementing SAR-SSNR. For Gaussian random Fourier feature mapping, we set the standard deviation σ of Gaussian distribution as 4 for 64 projection views of the clinical dataset and 5 for other cases. Since the original paper only experimented with 128 and 64 projection views, we simply selected $\sigma = 5$ for 256 projection views of XCAT and clinical dataset. Note that SAR-SSNR took around 1.5 and 2 h per image to optimize the MLP for XCAT and clinical data, respectively. Therefore, due to difficulty in conducting SAR-SSNR on all test images used for the proposed method, we used one slice for each projection view and dataset to compare the performance of the proposed method.
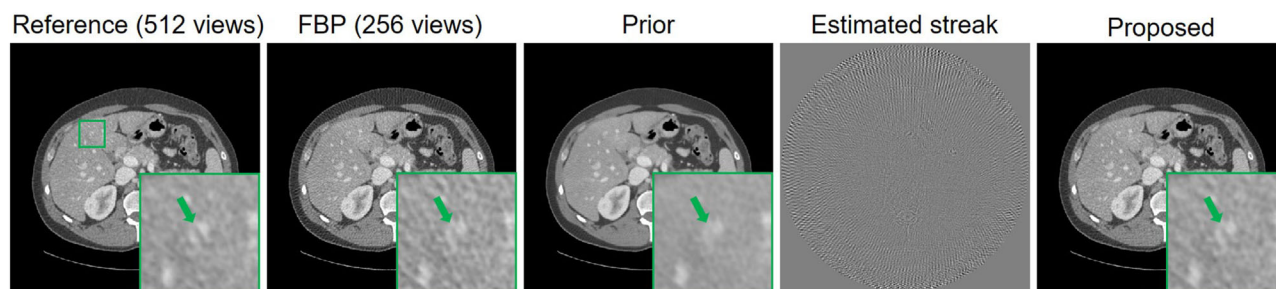
4 | RESULTS

4.1 | Qualitative results

Figure 4 presents the exampled images to describe the procedure of the proposed method, showing full-sampled reference, sparse-view FBP, generated prior, estimated streak artifacts, and the final corrected image by the proposed method. We show the enlarged ROIs using green boxes for reconstructed images in Figure 4. It can be observed that the streak artifacts are reduced in the prior image, whereas the tissue shape indicated by the green arrow was blurred. We estimated the original streak artifacts in FBP by utilizing the prior image, and its

output was shown as the result of the estimated streak artifacts. Then, the result of the proposed method was produced by subtracting the estimated streak artifacts from the FBP image. The proposed method preserved the edge sharpness of tissue structures similar to the reference image while reducing the streak artifacts compared to FBP.

Figures 5 and 6 show the resulting images of reference, FBP, TV-IR, FBPConvNet, AirNet, the proposed method, and equivalent-dose reference on the XCAT and clinical dataset, respectively. We also present the zoomed-in ROIs marked by green boxes in Figures 5 and 6. Note that the results of equivalent-dose reference are low-dose full-view CT images, where their dose levels are matched with the sparse-view CT images by considering the downsampling ratio between full-view and sparse-view projection views. The results of TV-IR contained a blocky appearance due to its TV term and the streak artifacts still remained in the reconstructed images. FBPConvNet and AirNet reduced the streak artifacts better than TV-IR by benefitting from fully-supervised learning. However, it can be seen that the fine details in the ROIs of FBPConvNet were blurred. For the ROIs of 64 projection views on the XCAT and clinical dataset, FBPConvNet removed some details overlapped by streak artifacts together and produced the distorted tissue shape as indicated by the green and yellow arrows. Since AirNet utilized the measured projection data for data fidelity term while integrating IR and DL framework, the overall results of AirNet showed better visual similarity to the reference image compared to those of FBPConvNet. However, AirNet still led to blurred fine details in some cases. In contrast, the proposed method effectively restored anatomical structures compared to other results for all cases.

However, the results of the proposed method contained high noise level because the proposed method reduced only streak artifacts. For fair comparison, we compared the results of the proposed method with those of equivalent-dose reference, where its total dose was set to be the same with sparse-view CT but acquired with full views. For the ROIs of XCAT dataset and 128 projection views of the clinical dataset, the proposed method preserved the noise texture similar to

FIGURE 4 Example of the intermediate results of the proposed method on clinical dataset with 256 projection views. The display window of the reconstructed images and difference image are [-240 320] in HU and [-100 100] in HU, respectively.
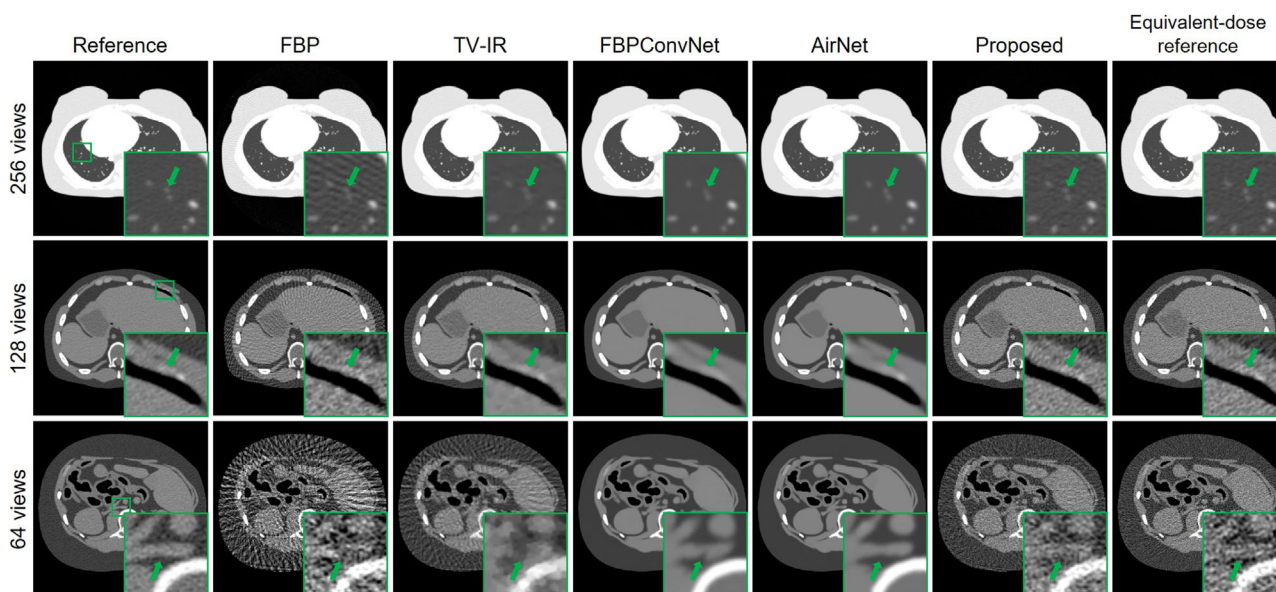
FIGURE 5 Reconstruction results on XCAT dataset under different projection views. The display window is [-1000 0] in HU for the first row and [-240 320] in HU for the second and third rows. XCAT, extended cardiac-torso.
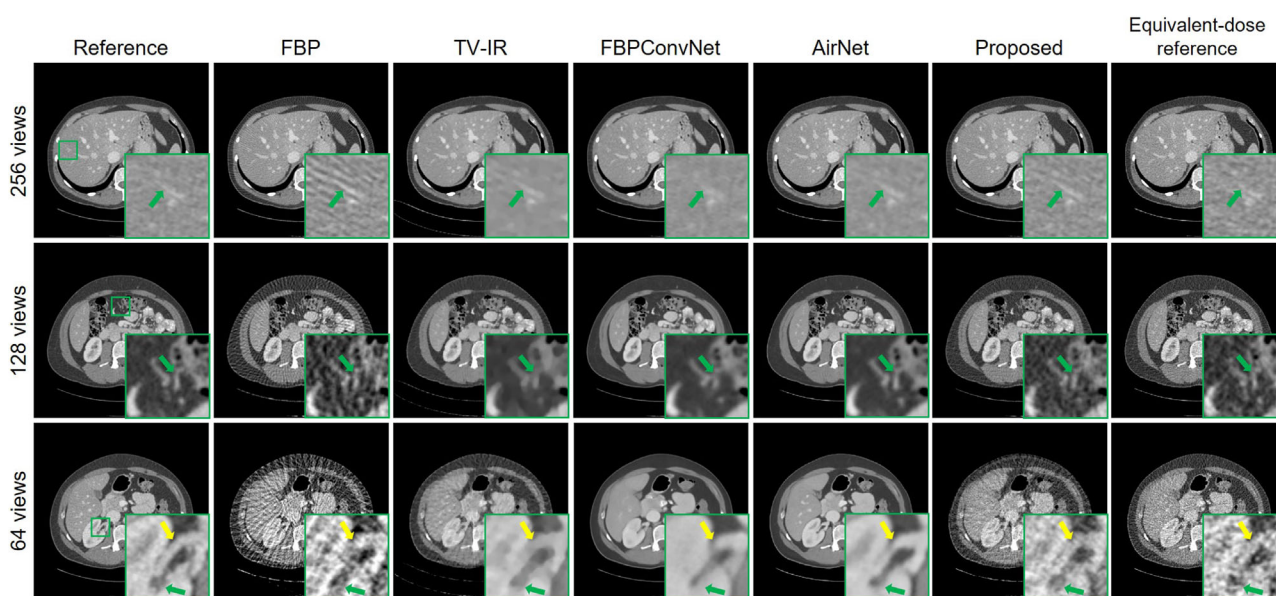
FIGURE 6 Reconstruction results on clinical dataset under different projection views. The display window is [-240 320] in HU.

TABLE 1 The quantitative evaluation of XCAT dataset under different projection views.

	256 views		128 views		64 views	
	NRMSE	SSIM	NRMSE	SSIM	NRMSE	SSIM
FBP	0.0137 ± 0.0018	0.8472 ± 0.0234	0.0286 ± 0.0036	0.6428 ± 0.0326	0.0539 ± 0.0073	0.5004 ± 0.0304
TV-IR	0.0095 ± 0.0023	0.9565 ± 0.0173	0.0121 ± 0.0029	0.9370 ± 0.0211	0.0189 ± 0.0048	0.8849 ± 0.0335
FBPConvNet	0.0061 ± 0.0013	0.9665 ± 0.0140	0.0067 ± 0.0013	0.9643 ± 0.0136	0.0088 ± 0.0016	0.9594 ± 0.0135
AirNet	$\mathbf{0.0059 \pm 0.0009}$	$\mathbf{0.9676 \pm 0.0130}$	$\mathbf{0.0064 \pm 0.0013}$	$\mathbf{0.9651 \pm 0.0135}$	$\mathbf{0.0069 \pm 0.0012}$	$\mathbf{0.9637 \pm 0.0135}$
Proposed	0.0098 ± 0.0022	0.9171 ± 0.0320	0.0145 ± 0.0024	0.8351 ± 0.0332	0.0254 ± 0.0024	0.6640 ± 0.0317

Note: The best values are indicated in boldface.
Abbreviations: NRMSE, normalized root mean-squared error; SSIM, the structural similarity index; XCAT, extended cardiac-torso.

TABLE 2 The quantitative evaluation of clinical dataset under different projection views.

	256 views		128 views		64 views	
	NRMSE	SSIM	NRMSE	SSIM	NRMSE	SSIM
FBP	0.0095 ± 0.0015	0.9251 ± 0.0190	0.0233 ± 0.0039	0.7303 ± 0.0425	0.0452 ± 0.0076	0.5288 ± 0.0485
TV-IR	0.0132 ± 0.0021	0.9585 ± 0.0137	0.0153 ± 0.0015	0.9380 ± 0.0100	0.0218 ± 0.0029	0.8828 ± 0.0231
FBPConvNet	0.0042 ± 0.0003	0.9853 ± 0.0018	0.0070 ± 0.0006	0.9686 ± 0.0037	0.0110 ± 0.0013	0.9464 ± 0.0073
AirNet	$\mathbf{0.0040 \pm 0.0004}$	$\mathbf{0.9864 \pm 0.0014}$	$\mathbf{0.0067 \pm 0.0005}$	$\mathbf{0.9717 \pm 0.0030}$	$\mathbf{0.0100 \pm 0.0010}$	$\mathbf{0.9551 \pm 0.0052}$
Proposed	0.0052 ± 0.0003	0.9772 ± 0.0025	0.0124 ± 0.0011	0.8953 ± 0.0125	0.0246 ± 0.0023	0.7282 ± 0.0261

Note: The best values are indicated in boldface.
Abbreviations: NRMSE, normalized root mean-squared error; SSIM, the structural similarity index.

equivalent-dose reference. However, for the ROIs of 256 projection views of clinical dataset, equivalent-dose reference did not restore the edge of the tissue indicated by the green arrow as accurately as the reference image due to its increased noise level. In contrast, the proposed method showed higher visual similarity of that detail to the reference image compared to equivalent-dose reference.

Although the proposed method was effective at preserving anatomical structures, it produced suboptimal imaging performance when the number of projection views was very low (e.g., 64 projection views). Because sparsier-view and sparse-view CT images were already significantly corrupted by streak artifacts, the trained CNN had difficulty in capturing useful image information of anatomical structures. Therefore, the generated prior images may not be as accurate as the reference image, resulting in relatively low performance in estimating the original streak artifacts.

4.2 | Quantitative results

We summarized the quantitative results of NRMSE and SSIM scores with the average and standard deviation in Tables 1 and 2 for XCAT and clinical dataset, respectively. Since the NRSME and SSIM scores are sensitive to noise, the conventional fully-supervised methods (i.e., FBPConvNet and AirNet) showed better scores despite over-smoothing than the proposed

method. As the results of the proposed method still had much remaining CT noise as the number of projection views decreases, the difference between NRMSE and SSIM values of the proposed method and those of CNN-based methods (i.e., FBPConvNet and AirNet) gradually increased.

4.3 | MTF comparison

Figure 7 presents MTF curves of reference, TV-IR, FBPConvNet, AirNet, and the proposed method. It can be observed that AirNet showed higher image resolution than FBPConvNet by utilizing measured projection data as a data fidelity term. In contrast, the proposed method provided the closest image resolution to reference image for all projection views, whereas the other results had a reduced resolution. Although the proposed method had more remaining errors as the number of projection views decreases, it can consistently preserve the image resolution similar to the reference image.

To further evaluate the resolution preservation of each method, we also calculated the full width at half maximum (FWHM) and the full width at tenth maximum (FWTM) of MTF. Table 3 summarizes FWHM and FWTM for each method under different projection views. Among the methods, the proposed method produced the highest FWHM and FWTM for all projection views, demonstrating its superior ability to preserve image resolution.
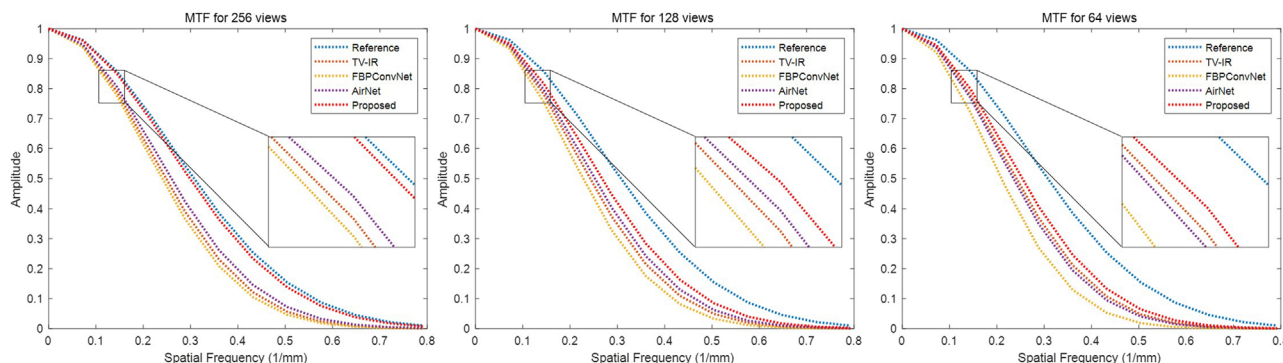
FIGURE 7 MTF results of different streak artifacts reduction methods under different projection views. MTF, modulation transfer function.

TABLE 3 FWHM and FWTM of MTF results under different projection views.

	256 views		128 views		64 views	
	FWHM	FWTM	FWHM	FWTM	FWHM	FWTM
TV-IR	0.2485	0.4554	0.2428	0.4431	0.2417	0.4405
FBPConvNet	0.2405	0.4376	0.2280	0.4178	0.2097	0.3875
AirNet	0.2597	0.4766	0.2526	0.4636	0.2352	0.4280
Proposed	**0.2987**	**0.5475**	**0.2676**	**0.4899**	**0.2540**	**0.4663**
Reference	0.3076	0.5621	0.3076	0.5621	0.3076	0.5621

Note: The best values are indicated in boldface.

Abbreviations: FWHM, full width at half maximum; FWTM, full width at tenth maximum; MTF, modulation transfer function.

4.4 | Comparison with self-supervised method

Figure 8 shows the results of the proposed method and SAR-SSNR on XCAT and clinical data. It can be observed that the proposed method and SAR-SSNR recovered the anatomical structures that closely resemble equivalent-dose reference. This demonstrates the proposed method's effectiveness in utilizing CNN to produce a prior image. Although both the image-based proposed method and sinogram-based SAR-SSNR effectively preserved the finer details, a fair comparison of their performances based on image quality is difficult due to the significantly longer training time required by SAR-SSNR.

4.5 | The performance of iterative prior generation

To investigate the effect of the number of iterations in obtaining prior images, we generated the prior images with iterations up to five and produced the final results using each prior image. Figure 9 visually compares the results of the proposed method obtained from different iterations with the corresponding difference images. For the ROIs of both 256 and 128 projection views, the performance of the proposed method was not significantly

different as the number of iterations increases. For 128 projection views, it can be seen that the difference images reduced the edge errors gradually as the number of iterations increases; however, those errors were saturated after two iterations. To quantitatively evaluate the quality of the resulting images, we plotted NRMSE and SSIM values in Figure 10. The results showed that the values of NRMSE and SSIM were the best at one iteration and two iterations for 256 and 128 projection views, respectively. This confirms that our calculation of the required number of iterations was appropriate.

## 4.6 | The performance of using noiseless XCAT dataset

To evaluate streak artifacts reduction performance of the proposed method without the effect of noise, we conducted experiments using noiseless XCAT dataset. For the proposed method, we used a downsampling ratio of 2 to generate noiseless sparsier-view CT images using noiseless projection data. While we conducted all procedures in the same manner as in the proposed method, we trained CNN by taking noiseless sparsier-view and sparse-view CT images as the input and target, respectively. After CNN training, we iteratively applied the trained CNN to noiseless sparse-view CT images until the strength of reduced streak artifacts
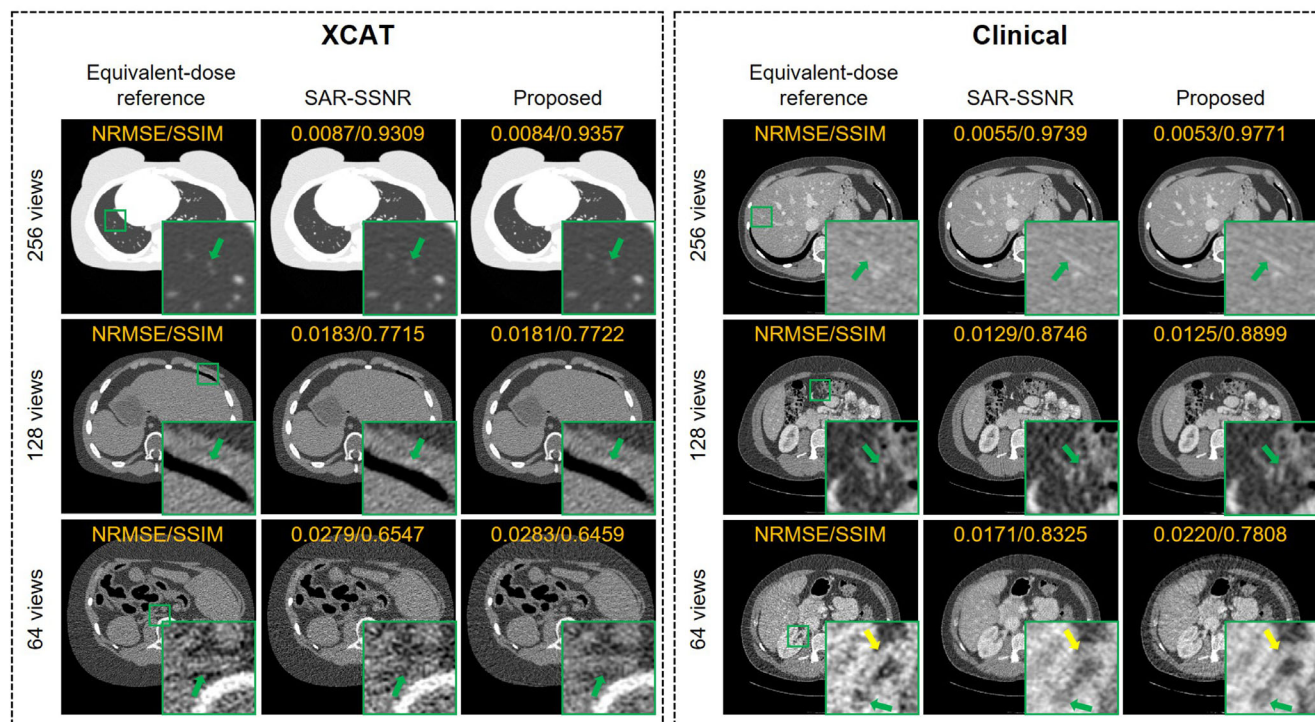
FIGURE 8 Reconstruction results of self-supervised-based methods on XCAT and clinical data under different projection views. The display window is [-1000 0] in HU for 256 projection views of XCAT dataset and [-240 320] in HU for other cases. XCAT, extended cardiac-torso.
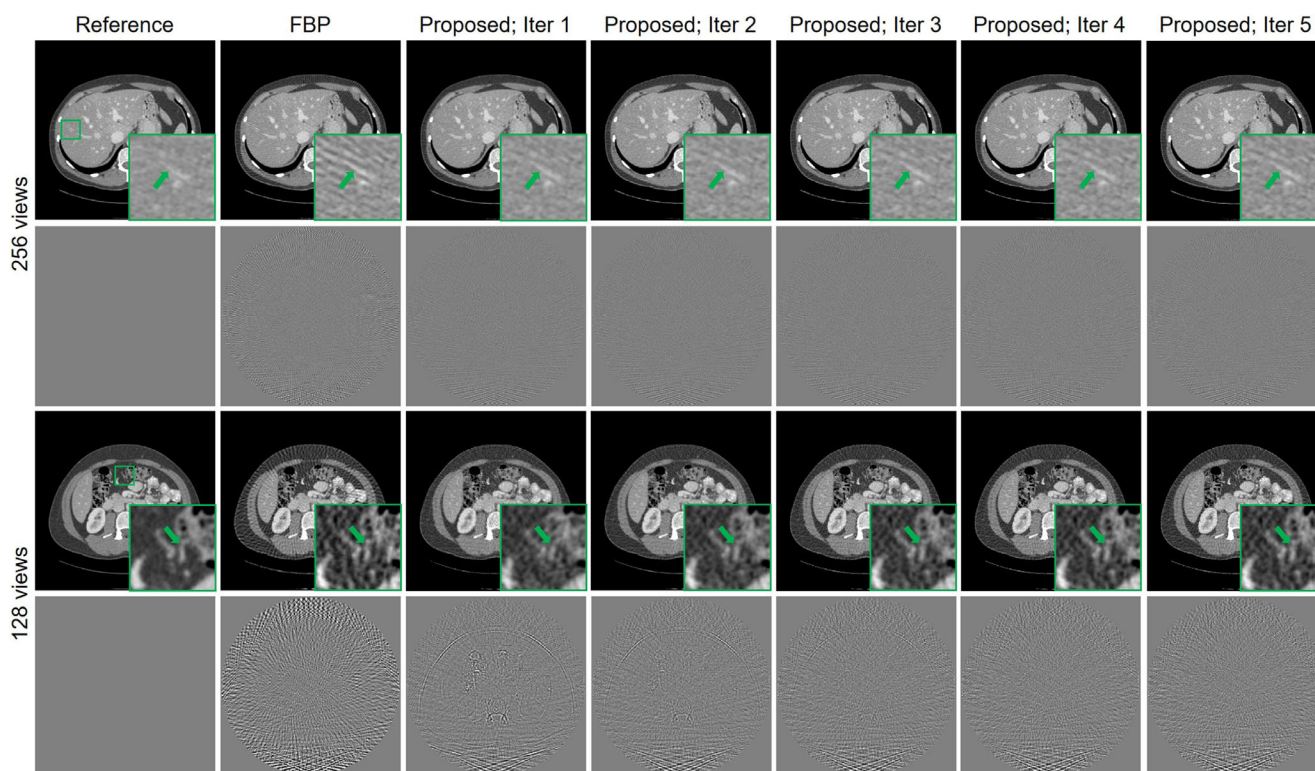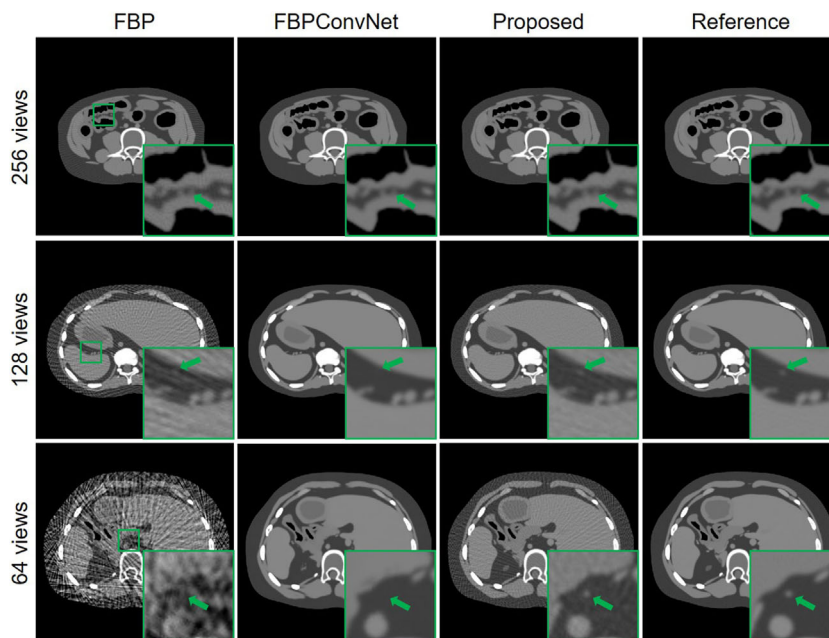

FIGURE 9 The results of the proposed method with different iterations of applying the trained network on clinical dataset under 256 and 128 projection views. The display window of the reconstructed images and difference images are [-240 320] in HU and [-200 200] in HU, respectively.
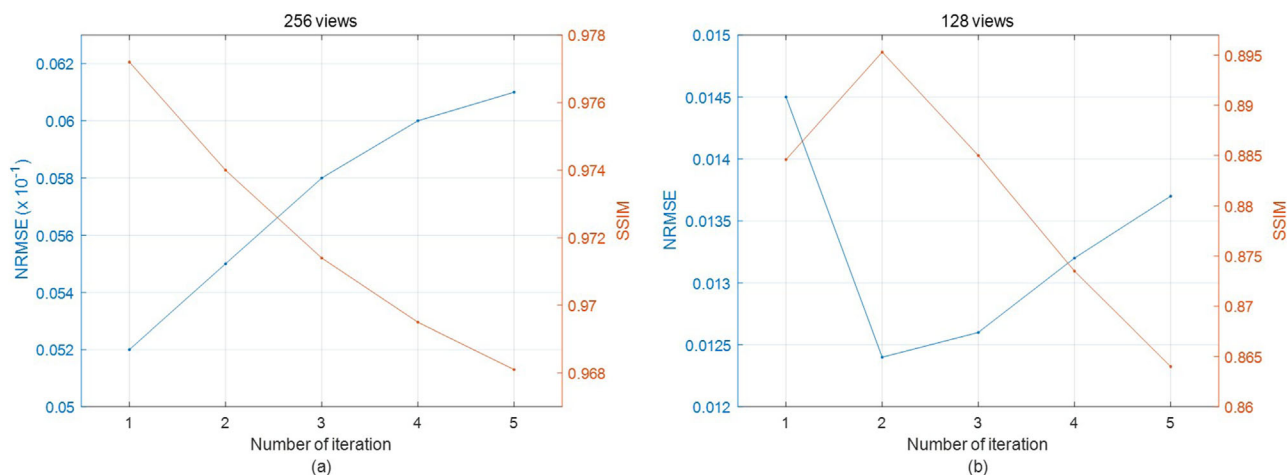
FIGURE 10 The NRMSE (blue) and SSIM (orange) scores of the proposed method for different iterations of applying the trained network on clinical dataset. (a) 256 projection views and (b) 128 projection views. NRMSE, normalized root mean-squared error; SSIM, the structural similarity index.

FIGURE 11 The results of FBPConvNet and the proposed method on noiseless XCAT dataset under different projection views. The display window is [-240 320] in HU. XCAT, extended cardiac-torso.

reaches to that of noiseless full-view CT images. For comparison, FBPConvNet was trained by taking noiseless sparse-view and full-view CT images as the input and target, respectively.

Figure 11 shows the results of FBP, FBPConvNet, the proposed method, and the reference image under different projection views. It can be observed that both FBPConvNet and the proposed method effectively reduced the streak artifacts that were noticeable in FBP. However, for the ROIs of 128 and 64 projection views, it becomes evident that the anatomical structures indicated by the green arrows were lost in FBPConvNet. In contrast, the proposed method successfully restored the original features, which were previously difficult to distinguish from streak artifacts or even invisible due to overlapping streak artifacts.

4.7 | The performance of using ultra-sparse views

The performance of the proposed method can be affected by the strength of the streak artifacts since the sparsier-view and sparse-view CT images used for CNN training have already been corrupted by streak artifacts. Therefore, we experimentally investigated the performance of the proposed method by reducing the number of projection views down to ultra-sparse views
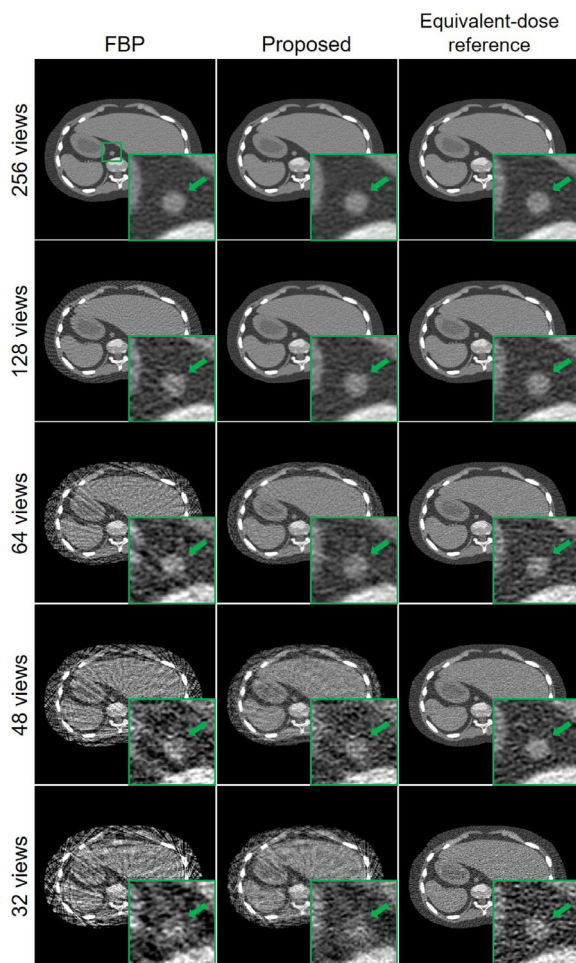
FIGURE 12 Visual comparison of FBP, the proposed method, and equivalent-dose reference image on XCAT dataset under various projection views. The display window is [-240 320] in HU. XCAT, extended cardiac-torso; FBP, filtered backprojection.

(i.e., 48 and 32 views) to determine the limitation of the proposed method in reducing streak artifacts. Note that since the required number of iterations (i.e., n) for prior image generation was not an integer on 48 views, we empirically set n to 3 based on the results of the proposed method which showed the lowest values on quantitative metrics.

Figure 12 shows the results of the proposed method as the number of projection views decreases. Until 48 projection views, the proposed method preserved the anatomical structures indicated by green arrows similar to equivalent-dose reference while reducing streak artifacts compared to FBP. However, for the ROIs of 32 projection views, the edge sharpness of the details was distorted, and the streak artifacts were not well reduced by the proposed method. Although the sparsier-view CT images corresponding to FBP images (i.e., sparse-view CT images) were not presented in Figure 12, the streak artifacts were very severe and fully covered both sparsier-view and sparse-view CT images used for CNN training on 32 projection views. Therefore, CNN would

have considered these images as containing only streak artifacts, which makes it difficult to capture fine details to be recovered in training. Nonetheless, it should be noted that the case of 32 projection views is a highly ill-posed scenario that makes most of the features invisible in the reconstructed images. Thus, the proposed image-based learning method has limitations in retrieving the details.

5 | DISCUSSION AND CONCLUSION

In this study, we proposed a self-supervised CNN approach to reduce streak artifacts in sparse-view CT images. To establish the training dataset with given unpaired sparse-view CT data, we downsampled given sparse-view sinogram and reconstructed sparsier-view CT images, which contained more severe streak artifacts than the original sparse-view CT images. After that, we trained CNN by taking sparsier-view and sparse-view CT images as the input and target, respectively. After CNN training, we obtained prior images by iteratively applying the trained network to given sparse-view CT images. For our final results, we estimated the original streak artifacts using prior images and subtracted the estimated streak artifacts from given sparse-view CT images, which was the key aspect of the proposed method.

To validate the performance of the proposed method, we compared it with several streak artifacts reduction methods. FBPConvNet and AirNet based on fully-supervised learning generally reduced both streak artifacts and CT noise, but in fact, it was at the expense of introducing severe image over-smoothing on anatomical structures. FBPConvNet distorted or blurred tissue shape as the number of projection views decreases (e.g., at 64 projection views). When the details were overlapped by strong streak artifacts, it is difficult for the network to distinguish the original structures from streak artifacts. Although AirNet, which utilizes measured projection data as a fidelity term, preserved tissue structures better than FBPConvNet, AirNet also produced blurred structures due to MSE loss used for CNN training. In contrast, from the qualitative and MTF results, the proposed method showed higher imaging performance in preserving fine details and image resolution compared to other methods.

While both the proposed method and SAR-SSNR are based on the self-supervised approach, it is important to note that there is a significant difference in the processing time. Since SAR-SSNR requires a test-time optimization, it takes 1.5 to 2 h to obtain a single image output. On the contrary, once the model is trained (6.5 h), the proposed method can produce results within 20 ms per image. Although optimizing a neural representation slice by slice offers the advantage of being patient-specific, its time-consuming nature remains a critical limitation. Therefore, the proposed

method which enables real-time application, is more clinically practical than SAR-SSNR.

Although the proposed method showed suboptimal performance on 64 projection views, it reduced streak artifacts and preserved anatomical structures better than those of FBP images. To further improve the performance of the proposed method, we can reuse the final output of the proposed method as the target of the network in the training phase and then conduct the same procedures of the proposed method. From this, since the network is trained to predict the target images with enhanced quality, high-quality prior images can be generated. Then, more accurate streak artifacts would be estimated, which can further increase the performance of the proposed method. The purpose of iteratively performing the proposed method is to make prior images similar to the reference images. Thus, our framework can be applied repeatedly (e.g., about two or three times) until the final results reduced the streak artifacts effectively.

Since the streak artifacts are strengthened as the number of projection views decreases, it is challenging to acquire the prior image as accurately as the reference image. To further improve the quality of the prior image, our framework can be extended into a dual-domain, as the success of dual-domain approach in sparse-view CT.[8,20,21] The first possible way is that in the training phase, we can train sinogram-domain CNN (e.g., enhanced deep super-resolution network [EDSR][40]) to interpolate sparsier-view sinograms into sparse-view sinograms and then reconstruct pseudo sparse-view CT images from the output of sinogram-domain CNN using FBP algorithm. After that, we train image-domain CNN with end-to-end manner to refine pseudo sparse-view CT images to given sparse-view CT images. In the testing phase, we acquire prior image by applying the trained CNN (interpolation for sinogram-domain and refinement for image-domain) to the given sparse-view sinogram iteratively until the interpolated sparse-view sinogram has the same size with full-view one. The second possible way is that after acquiring prior images conducted in the same manner as in the proposed method, we first generate corresponding sparse-view sinograms by forward projection of prior images. We then train sinogram-domain CNN from sparse-view sinogram of prior image to given sparse-view sinogram to ensure consistency. These approaches can fully utilize the measured data, enabling us to overcome the limitation of self-supervised methods on even much sparsier projection views. This will be investigated in future work.

In our experiment, we set the downsampling ratio as 2 to generate sparsier-view CT images. However, this downsampling ratio of 2 used for training dataset generation was a challenging condition in the training phase because sparsier-view CT images contained only half of the information of the measured data (i.e., given

sparse-view sinogram). Nonetheless, the proposed method preserved fine details better than the other methods. The important component of our framework for improving the imaging performance is to generate prior images as accurately as the reference images, which directly leads to increased accuracy of streak artifacts estimation. For better quality of prior images, the downsampling ratio to generate sparsier-view CT images can be chosen smaller than 2 (e.g., 1.2 or 1.5). When the small downsampling ratio is used, each iteration reduces less amount of streak artifacts and preserves more fine details, facilitating streak artifacts reduction task by treating it as several simple sub-tasks. This would be investigated in future work.

Since the proposed method was designed to reduce only streak artifacts, the CT noise still remained in the results of the proposed method. To reduce the remaining CT noise in the results of the proposed method, the denoising processing (e.g., BM3D[41]) which does not require full-view CT data can be applied after our framework is performed. Since the performance of those methods depends on controllable parameters (e.g., sigma value for BM3D), we need to carefully optimize those parameters while keeping the ability of structure preservation of the proposed method. This would be researched in future work.

In this work, we used simple U-Net-based architecture and MSE loss for CNN training. However, several advanced ones had been proposed in the medical imaging field such as tight-frame U-Net,[5] DenseNet with deconvolution,[6] and attention-based U-Net[42,43] for the network architecture and perceptual[44,45] and observer[46,47] losses for the loss function. The network architecture and loss function in our framework can be replaced with those components. However, this requires further investigation of which combinations can improve the performance of the proposed method. Since this study focused on developing the new framework rather than improving network architecture or loss function, this experiment would be researched in future work.

As the dataset with three-dimensional volumes of cone-beam geometry requires high memory usage, we conducted our framework using two-dimensional CT images generated on fan-beam geometry due to our limited memory resources. However, the proposed method can be extended to cone-beam CT (CBCT). To regulate memory requirements within the allowed system, we can group the adjacent slices (e.g., approximately 10 slices) of a dataset for CNN training. As the reconstructed images generated by CBCT are correlated over image slices, the information of the streak artifacts would be shared across adjacent slices. Therefore, the network can be trained to maintain useful interslice information by learning not only distributed patterns but also the spatial correlation of the streak artifacts between grouped slices. We expect this approach can further improve imaging performance.

In conclusion, we proposed a novel framework for streak artifacts reduction in sparse-view CT using only unpaired sparse-view CT data. Although we did not use any information from full-view CT data, we were able to recover the finer structures by utilizing CNN-based prior images for streak artifacts estimation and subtracting those additive errors from the FBP images. From the results, the proposed method showed the highest imaging performance compared to the other results while preserving the image resolution. We expect that our framework can overcome the difficulty of paired data acquisition and be utilized in the medical imaging field.

ACKNOWLEDGMENTS

CONFLICTS OF INTEREST STATEMENT
The authors have no conflicts to disclose.

DATA AVAILABILITY STATEMENT
The data that support the findings of this study are available from the corresponding author upon reasonable request.

REFERENCES
1. Hsieh J. Computed tomography: principles, design, artifacts, and recent advances; vol. 114. SPIE Press; 2003 2003.
2. Kim B, Han M, Shim H, Baek J. A performance comparison of convolutional neural network-based image denoising methods: the effect of loss functions on low-dose CT images. *Med Phys*. 2019;46(9):3906-3923.
3. Zhang Q, Hu Z, Jiang C, Zheng H, Ge Y, Liang D. Artifact removal using a hybrid-domain convolutional neural network for limited-angle computed tomography imaging. *Phys Med Biol*. 2020;65(15):155010.
4. Hu Chen, Yi Zhang, Yunjin Chen, et al. Learn: learned experts' assessment-based reconstruction network for sparse-data CT. *IEEE Trans Med Imaging*. 2018;37(6):1333-1347.
5. Han Y, Ye JC. Framing U-Net via deep convolutional framelets: application to sparse-view CT. *IEEE Trans Med Imaging*. 2018;37(6):1418-1429.
6. Zhang Z, Liang X, Dong X, Xie Y, Cao G. A sparse-view CT reconstruction method based on combination of DenseNet and deconvolution. *IEEE Trans Med Imaging*. 2018;37(6):1407-1417.
7. Jin KH, McCann MT, Froustey E, Unser M. Deep convolutional neural network for inverse problems in imaging. *IEEE Trans Image Process*. 2017;26(9):4509-4522.
8. Wu W, Hu D, Niu C, Yu H, Vardhanabhuti V, Wang G. Drone: dual-domain residual-based optimization network for sparse-view CT reconstruction. *IEEE Trans Med Imaging*. 2021;40(11):3002-3014.
9. Zhang C, Li Y, Chen GH. Accurate and robust sparse-view angle CT image reconstruction using deep learning and prior image constrained compressed sensing (DL-PICCS). *Med Phys*. 2021;48(10):5765-5781.
10. Bian J, Siewerdsen JH, Han X, et al. Evaluation of sparse-view reconstruction from flat-panel-detector cone-beam CT. *Phys Med Biol*. 2010;55(22):6575.
11. Park JC, Song B, Kim JS, et al. Fast compressed sensing-based CBCT reconstruction using Barzilai-Borwein formulation for application to on-line IGRT. *Med Phys*. 2012;39(3):1207-1217.
12. Sidky EY, Pan X. Image reconstruction in circular cone-beam computed tomography by constrained, total-variation minimization. *Phys Med Biol*. 2008;53(17):4777.
13. Liu Y, Ma J, Fan Y, Liang Z. Adaptive-weighted total variation minimization for sparse data toward low-dose X-ray computed tomography image reconstruction. *Phys Med Biol*. 2012;57(23):7923.
14. Tian Z, Jia X, Yuan K, Pan T, Jiang SB. Low-dose CT reconstruction via edge-preserving total variation regularization. *Phys Med Biol*. 2011;56(18):5949.
15. Kim B, Shim H, Baek J. Weakly-supervised progressive denoising with unpaired CT images. *Medical Image Analysis*. 2021;71:102065.
16. Liang K, Yang H, 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.
17. Zheng A, Gao H, Zhang L, Xing Y. A dual-domain deep learning-based reconstruction method for fully 3D sparse data helical CT. *Phys Med Biol*. 2020;65(24):245030.
18. Lee D, Choi S, Kim HJ. High quality imaging from sparsely sampled computed tomography data with deep learning and wavelet transform in various domains. *Med Phys*. 2019;46(1):104-115.
19. Chen G, Hong X, Ding Q, et al. Airnet: fused analytical and iterative reconstruction with deep neural network regularization for sparse-data CT. *Med Phys*. 2020;47(7):2916-2930.
20. Wu W, Guo X, Chen Y, Wang S, Chen J. Deep embedding-attention-refinement for sparse-view CT reconstruction. *IEEE Trans Instrum Meas*. 2022;72:1-11.
21. Pan J, Zhang H, Wu W, Gao Z, Wu W. Multi-domain integrative swin transformer network for sparse-view tomographic reconstruction. *Patterns*. 2022;3(6):100498.
22. Lewis MA, Pascoal A, Keevil SF, Lewis CA. Selecting a CT scanner for cardiac imaging: the heart of the matter. *Brit J Radiol*. 2016;89(1065):20160376.
23. Kudo H, Suzuki T, Rashed EA. Image reconstruction for sparse-view CT and interior CT–introduction to compressed sensing and differentiated backprojection. *Quant Imaging Med Surg*. 2013;3(3):147.
24. Su T, Cui Z, Yang J, et al. Generalized deep iterative reconstruction for sparse-view CT imaging. *Phys Med Biol*. 2022;67(2):025005.
25. Pearce MS, Salotti JA, Little MP, et al. Radiation exposure from CT scans in childhood and subsequent risk of leukaemia and brain tumours: a retrospective cohort study. *Lancet*. 2012;380(9840):499-505.
26. Moran N, Schmidt D, Zhong Y, Coady P. Noisier2noise: learning to denoise from unpaired noisy data. In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. IEEE; 2020:12064-12072.
27. Imai K, Ikeda M, Enchi Y, Niimi T. Statistical characteristics of streak artifacts on CT images: relationship between streak artifacts and mA s values. *Med Phys*. 2009;36(2):492-499.
28. Imai K, Ikeda M, Enchi Y, Niimi T. A detection method for streak artifacts and radiological noise in a non-uniform region in a CT image. *Physica Med*. 2010;26(3):157-165.
29. Kim S, Kim B, Baek J. Sparsier2sparse: weakly supervised learning for streak artifact reduction with unpaired sparse-view CT data. In: *7th International Conference on Image Formation in X-Ray Computed Tomography*. SPIE; 2022;12304:124-129.

30. Kim B, Shim H, Baek J. A streak artifact reduction algorithm in sparse-view CT using a self-supervised neural representation. *Med Phys*. 2022;49(12):7497-7515.

31. Kak AC, Slaney M. *Principles of Computerized Tomographic Imaging*. SIAM; 2001.

32. Segars WP, Mahesh M, Beck TJ, Frey EC, Tsui BMW. Realistic CT simulation using the 4D XCAT phantom. *Med Phys*. 2008;35(8):3800-3808.

33. McCollough CH, Bartley AC, Carter RE, Chen B, et al. Low-dose CT for the detection and classification of metastatic liver lesions: results of the 2016 low dose CT grand challenge. *Med Phys*. 2017;44(10):e339-e352.

34. Siddon RL. Fast calculation of the exact radiological path for a three-dimensional CT array. *Med Phys*. 1985;12(2):252-255.

35. Kim S, Ahn J, Kim B, Kim C, Baek J. Convolutional neural network–based metal and streak artifacts reduction in dental CT images with sparse-view sampling scheme. *Med Phys*. 2022;49(9):6253-6277.

36. Odena A, Dumoulin V, Olah C. Deconvolution and checkerboard artifacts. *Distill*. 2016;1(10):e3.

37. Kingma DP, Ba J. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*. 2014.

38. Wang Z, Bovik AC, Sheikh HR, Simoncelli EP. Image quality assessment: from error visibility to structural similarity. *IEEE Trans Image Process*. 2004;13(4):600-612.

39. Chen B, Christianson O, Wilson JM, Samei E. Assessment of volumetric noise and resolution performance for linear and nonlinear CT reconstruction methods. *Med Phys*. 2014;41(7):071909.

40. Lim B, Son S, Kim H, Nah S, Lee KM. Enhanced deep residual networks for single image super-resolution. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*. IEEE; 2017:136-144.

41. Danielyan A, Katkovnik V, Egiazarian K. Bm3D frames and variational image deblurring. *IEEE Trans Image Process*. 2011;21(4):1715-1728.

42. Oktay O, Schlemper J, Folgoc LL, et al. Attention U-Net: learning where to look for the pancreas. *arXiv preprint arXiv:1804.03999*. 2018.

43. Li W, Qin S, Li F, Wang L. Mad-unet: a deep u-shaped network combined with an attention mechanism for pancreas segmentation in CT images. *Med Phys*. 2021;48(1):329-341.

44. Johnson J, Alahi A, Fei-Fei L. Perceptual losses for real-time style transfer and super-resolution. In: *European Conference on Computer Vision*. Springer; 2016:694-711.

45. Han M, Shim H, Baek J. Perceptual CT loss: implementing CT image specific perceptual loss for cnn-based low-dose CT denoiser. *IEEE Access*. 2022;10:62412-62422.

46. Han M, Shim H, Baek J. Low-dose CT denoising via convolutional neural network with an observer loss function. *Med Phys*. 2021;48(10):5727-5742.

47. Kim G, Han M, Shim H, Baek J. A convolutional neural network-based model observer for breast CT images. *Med Phys*. 2020;47(4):1619-1632.

48. Gumbel EJ. Statistics of extremes. New York: Columbia University Press; 1958.

## APPENDIX A: STREAK ARTIFACTS ANALYSIS USING GUMBEL EVALUATION METHOD

We investigated whether the streak artifacts caused by sparse view can be modeled by Gumbel distribution. In our framework, we can consider two types of streak artifacts: (1) those generated by subtracting I_{full} from I_{sparse} (i.e., $I_{sparse} - I_{full}$), and (2) those generated by subtracting I_{sparse} from $I_{sparsier}$ (i.e., $I_{sparsier} - I_{sparse}$), as shown in Figure A1.

To show the statistical characteristic of streak artifacts, we followed the overall procedures of the Gumbel evaluation method conducted in Imai et al.[28] For streak artifacts analysis, we used ROI with a size of 80 by 100 pixels as indicated by red box in Figure A1 and measured CT value profile in a direction parallel to 80 pixels (line "a-b" within ROI). Since CT values of streak artifacts are considered higher than those of CT noise, we obtained the maximum CT value for each line profile by traversing along the white arrow in ROI with a one-pixel interval.[28] After obtaining the 100 largest CT values, we utilized them to estimate the cumulative probability function of Gumbel distribution based on the mean rank method with order statistics.[48] We applied this Gumbel evaluation method to both $I_{sparse} - I_{full}$ and $I_{sparsier} - I_{sparse}$, of which cumulative probability functions are denoted as $f(I_{sparse}|I_{full})$ and $f(I_{sparsier}|I_{sparse})$, respectively.

Figure A2 shows the distribution of the largest CT values with corresponding estimated cumulative probability for $I_{sparse} - I_{full}$ and $I_{sparsier} - I_{sparse}$, respectively. We also visualized their overlapped ones for comparison as well. It can be observed that the CT-value variation of streak artifacts increases as the number of projection views decreases. Since the largest CT values are
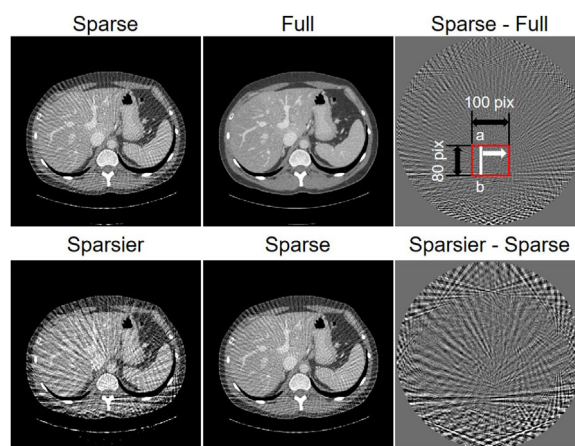
FIGURE A1 Examples of streak artifacts on clinical dataset with 128 projection views for Gumbel distribution modeling analysis. The first row shows streak artifacts between sparse-view and full-view CT images. The second row shows streak artifacts between sparsier-view and sparse-view CT images. The display window is [-240 320] in HU. CT, computed tomography.
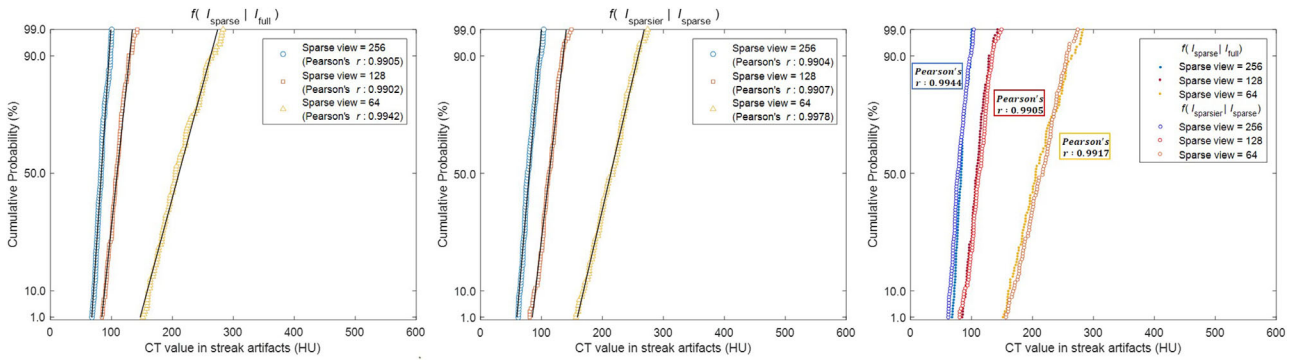
FIGURE A2 Gumbel plots for two types of streak artifacts under different sparse projection views.

linearly distributed on Gumbel probability (Pearson's coefficients ≥ 0.99) for all cases, we can confirm that both $f(I_{sparse}|I_{full})$ and $f(I_{sparsier}|I_{sparse})$ can be modeled by Gumbel distribution. This indicates that the statistical characteristic of the streak artifacts can be described using this distribution. Furthermore, for the overlapped plots between $f(I_{sparse}|I_{full})$ and $f(I_{sparsier}|I_{sparse})$, it can be seen that those distributions are nearly identical (Pearson's coefficients ≥ 0.99). This demonstrates that the distributions of the training dataset (i.e., $I_{sparsier}$ and I_{sparse}) cover those of the test dataset (i.e., I_{sparse} and I_{full}), indicating the feasibility of reducing streak artifacts of the proposed method.

APPENDIX B: ANALYSIS OF THE UPPER BOUND PERFORMANCE OF THE PROPOSED METHOD

The degradation of quantitative results of the proposed method was larger than that of other methods as the number of projection views decreases. This is mainly because the proposed method still contains CT noise, unlike other methods. Although it is difficult to present the strict accuracy of the proposed method due to inherently heuristic property of DL algorithms, we derive that the upper bound performance of the proposed method would be an equivalent-dose reference under mild conditions.

The reconstruction for CT image can be formulated as linear system:

$$s = A\mu + n \tag{B1}$$

where s is measured log-normalized sinogram (projection data), A is the matrix for CT image modeling system, μ is high-quality image to be acquired and n is projection noise. Then, the sparse-view and full-view CT images can be represented as follows:

$$I_{sparse} = F_{sparse}(A_{sparse}\mu + n) \tag{B2}$$

$$I_{full} = F_{full}(A_{full}\mu + n) \tag{B3}$$

where F_{sparse} and F_{full} are the matrixes of the FBP operator performed in sparse-view and full-view geometry, respectively. A_{sparse} and A_{full} are the matrixes for the image modeling system in sparse-view and full-vie geometry, respectively. Note that I_{sparse} and I_{full} had the same dose level for each projection view. Since FBP algorithm is a linear system, we can express (B2)-(B3) with two terms as follows:

$$I_{sparse} - I_{full} = (F_{sparse}A_{sparse}\mu - F_{full}A_{full}\mu)$$

$$+ (F_{sparse}n - F_{full}n) \tag{B4}$$

The first term is the original streak artifacts I_{streak}, which do not contain noise, and the second term is noise-induced image.

Let σ^2_{full} and k be the noise variance of I_{full} and the ratio between the number of full and sparse projection views (i.e., N_{full}/N_{sparse}), respectively. Note that the noise variance of $F_{full}n$ would also be σ^2_{full} because μ is noise-free image. Since the pixel values of the reconstructed images are averaged over their corresponding number of projection views, the noise level of the reconstructed images under sparse projection views would be larger than that of full projection views when the dose level per each projection view is fixed.[30] Therefore, the noise variance of $F_{sparse}n$ can be considered k times that of $F_{full}n$. To validate this, we calculated the noise variance both on $F_{sparse}n$ and $F_{full}n$. We first acquired noise-only projection data (i.e., n) by subtracting noise-free projection data (i.e., $A\mu$) from measured projection data (i.e., s) under sparse and full projection views, respectively. Then, we reconstructed noise-only images from noise-only projection data using FBP algorithm. Figure B1 shows the noise-only images and corresponding noise variance value. It can be observed that the noise variance of sparse-view noise-only image is about four times that of full-view one, confirming the inverse-proportional relationship between the noise variance and the number of projection views.
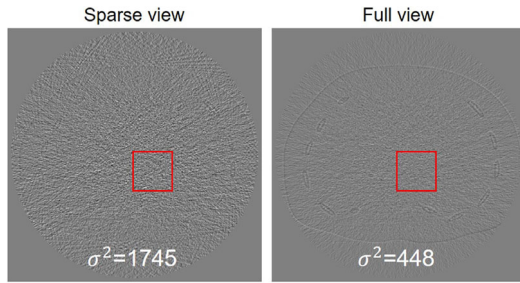
FIGURE B1 Example of the noise variance evaluation of noise-only images on XCAT dataset under 128 and 512 projection views, respectively. The values of noise variance are calculated on ROI indicated by red boxes. The display window is [-200 200] in HU. XCAT, extended cardiac-torso.

For the proposed method, the final output $I_{proposed}$ can be derived as follows:

$$
\begin{aligned}
I_{proposed} &= I_{sparse} - \tilde{I}_{streak} \\
&= I_{full} + I_{streak} + (F_{sparse}n - F_{full}n) - \tilde{I}_{streak} \\
&= (F_{full}A_{full}\mu + F_{full}n) \\
&\quad + (I_{streak} - \tilde{I}_{streak}) + (F_{sparse}n - F_{full}n) \\
&= F_{full}A_{full}\mu + (I_{streak} - \tilde{I}_{streak}) + F_{sparse}n
\end{aligned} \tag{B5}
$$

where $\tilde{I}_{streak}$ is the estimated streak artifacts using prior images. From the Equation (B5), the noise variance of

$I_{proposed}$ can be calculated as follows:

$$
\begin{aligned}
var(I_{proposed}) &= var(F_{full}A_{full}\mu + I_{streak} - \tilde{I}_{streak} + F_{sparse}n) \\
&= var(F_{sparse}n) \\
&= k \cdot \sigma^2_{full} = var(I_{equivalent-dose})
\end{aligned} \tag{B6}
$$

where the noise variance of $F_{full}A_{full}\mu$ and $I_{streak} - \tilde{I}_{streak}$ can be regarded as zero because these terms do not contain noise. $I_{equivalent-dose}$ are equivalent-dose references, where their dose levels are matched with I_{sparse}. Therefore, $I_{proposed}$ inevitably has the noise with the variance of $k \cdot \sigma^2_{full}$ while also containing the residual errors of streak artifacts. If the prior image is as accurate as reference, the estimated streak artifacts could be the same as the original ones, and thus the residual errors of streak artifacts would be zero. However, as the number of projection views decreases, it is difficult to acquire accurate prior images, and the residual errors of streak artifacts would increase. As a result, the performance of the proposed method could not outperform that of equivalent-dose reference. From this reason, the degradation of the quantitative results of the proposed method increases as the number of projection views decreases due to remaining noise and residual streak artifacts errors.