

Article

Self-Supervised Noise Reduction in Low-Dose Cone Beam Computed Tomography (CBCT) Using the Randomly Dropped Projection Strategy

Young-Joo Han ^{1,2}  and Ha-Jin Yu ^{2,*} ¹ R&D Center, Vieworks, Anyang-si 14055, Korea; orora71@gmail.com² School of Computer Science, University of Seoul, Seoul 02504, Korea

* Correspondence: hjyu@uos.ac.kr; Tel.: +82-2-6490-2448

Abstract: Deep learning-based denoising methods have proved efficient for medical imaging. Obtaining a three-dimensional representation of a scanned object is essential, such as in the computed tomography (CT) system. A sufficient radiation dose needs to be irradiated to a scanned object to obtain a high-quality image. However, the radiation dose is insufficient in many cases due to hardware limitations or health care issues. A deep learning-based denoising method can be a solution to obtaining good images, even when the radiation dose is insufficient. However, most existing deep learning-based denoising methods require numerous paired low-dose CT (LDCT) images and normal-dose CT (NDCT) images. It is almost impossible to obtain numerous well-paired LDCT and NDCT images. Self-supervised denoising methods were proposed to train a denoising neural network on only noisy images. These methods can be applied to the projection domain in LDCT. However, applying denoising in the projection image domain is a challenging task, because the projection images for LDCT have extremely weak signals. To solve this problem, we propose a noise reduction method based on the dropped projection strategy. The proposed method works by first reconstructing the 3D image with the degraded versions of the projection images generated by Bernoulli sampling. Subsequently, the denoising neural network is trained to restore the signal dropped out by Bernoulli sampling in the projection image domain. As such, the method we propose solves the over-smoothing problem in previous methods and is able to be trained with a small amount of data. We verified the performance of our proposed method on the SPARE challenge dataset and the in-house lithium polymer dataset. The experiments on two datasets show that the proposed method outperforms the conventional denoising methods by at least 4.47 dB of PSNR value.

Keywords: computed tomography (CT); denoising; noise reduction; self-supervised

Citation: Han, Y.-J.; Yu, H.-J. Self-Supervised Noise Reduction in Low-Dose Cone Beam Computed Tomography (CBCT) Using the Randomly Dropped Projection Strategy. *Appl. Sci.* **2022**, *12*, 1714. <https://doi.org/10.3390/app12031714>

Academic Editor: Vincent A. Cicirello

Received: 24 December 2021

Accepted: 4 February 2022

Published: 7 February 2022

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Computed tomography (CT) is an essential technique to obtain a three-dimensional (3D) representation of a scanned object in medical imaging, industrial imaging, and other areas. In medical imaging, CT can determine injuries or diseases in nearly all parts of the body. In industrial imaging, CT can be used for nondestructive testing to detect cracks or defects.

Irradiating a target object with a sufficient radiation dose is necessary to obtain a good quality 3D image. However, in real-world situations, a target object cannot be sufficiently irradiated due to hardware limitations or public healthcare issues. Especially in medical imaging, the radiation in CT can cause cancer and genetic or other diseases [1]. Thus, it is crucial to minimize and optimize the radiation dose in CT.

In addition, CT images are reconstructed from a plurality of the projections acquired from an X-ray detector as the X-ray tube and detector rotate through the target object. The radiation dose reduction causes severe photon noise with a Poisson distribution [2] in

the projection images, degrading the quality of the 3D reconstructed image. For example, photon noise in projection images is a source of streak artifacts in reconstructed images.

Numerous studies in low-dose CT (LDCT) have been proposed using either conventional image processing-based denoising methods [3–16] or deep learning-based denoising methods [17–22]. Conventional image processing-based denoising methods show high performance. However, there are some problems, such as a removed microstructure, blurred edges, or high computational complexity. To solve these problems, deep learning-based approaches are proposed. These methods have exhibited superior performance. However, there are still problems, namely that it is difficult to obtain appropriate datasets due to public health care issues (the details are described in Section 2). Therefore, the need for deep learning-based denoising methods using self-supervised or unsupervised training schemes has emerged.

Even in noise reduction in the natural image domain, it is challenging to obtain paired noisy and clean images for training a supervised denoising network. Several studies that train networks without paired noisy and clean images in the natural image domain have been proposed to handle this problem. The Noise2Noise [23] method has proved that denoising networks can be trained using paired noisy and noisy images in specific environments. However, even if the Noise2Noise approach can train networks without paired noisy and clean images, the method still requires paired noisy and noisy images with independent noise. Based on the Noise2Noise theory, several studies that can denoise an image without paired images have been proposed to address this problem.

Noise2Void (N2V) [24] introduces the blind spot strategy with self-prediction loss to train networks without paired images. The blind spot strategy allows for effectively predicting artificially missing pixels using the neighboring noisy pixels. Noise2Self (N2S) [25] also demonstrates training neural networks using self-supervision for noise reduction in the natural image domain. Motivated by these studies, the Self2Self [26] approach proposed a method of training networks with pairs of Bernoulli-sampled input images from a single image. The Self2Self approach uses the method with dropout of both the input image and convolutional layers in the decoding part and the method of estimating noise-reduced results by averaging predictions generated from multiple outputs of the trained model. These methods allow for lowering the variance of the denoising neural network, which can be interpreted as a Bayes estimator. The dropout-based ensemble method is very efficient and demonstrates superior performance in noise reduction without paired images.

Although these methods have exhibited superior performance in the natural image domain, they are challenging to apply to noise reduction in reconstructed LDCT images because of the propositions in the blind spot strategy. The methods using the blind spot strategy have two simple statistical propositions: (1) the signal is not pixel-wise independent, and (2) the noise is pixel-wise independent. The methods that use the blind spot strategy perform noise reduction correctly when satisfying these propositions.

Deep learning-based noise reduction approaches in LDCT can be divided into the following types: noise reduction in the reconstructed image domain and noise reduction in the projection or sinogram domain. The noise caused by an insufficient radiation dose is photon noise with a Poisson distribution. Photon noise can be treated as pixel-wise independent in projection images. However, the reconstructed image is created by reconstructing the signal of the plurality of the projection image. Thus, the noise in the reconstructed image is not pixel-wise independent. Because of this phenomenon, if a blind spot strategy-based noise reduction technique is performed on reconstructed data, artifacts caused by photon noise are recognized as structures, emphasizing them more. Figure 1 illustrates that artifacts are emphasized when the noise reduction technique with the blind spot strategy is performed in the reconstructed image domain.

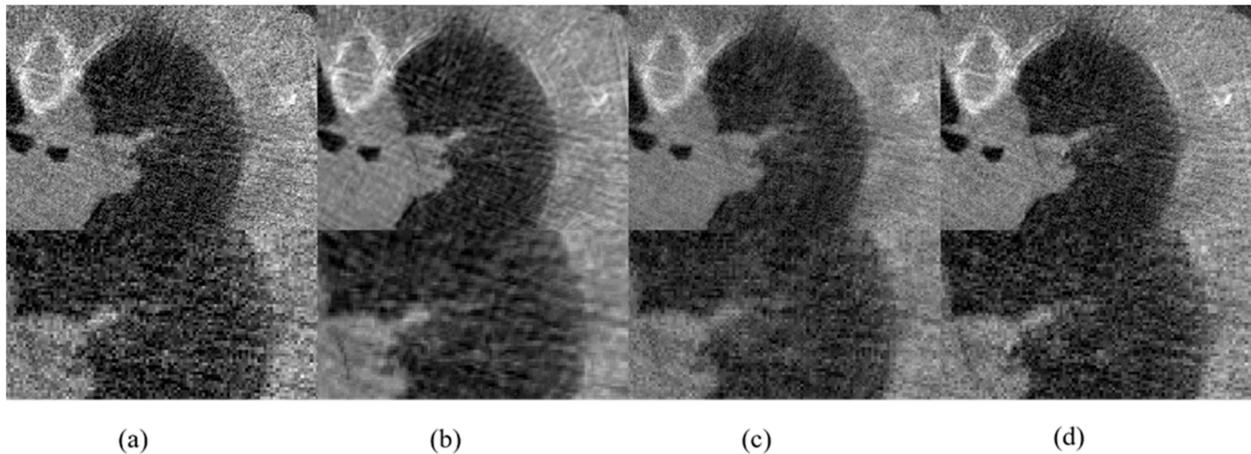


Figure 1. Visual comparison from the SPARE dataset: (a) input noisy image, (b) resulting image of denoising in the reconstructed image domain, (c) resulting image of our proposed method, and (d) target image.

For these reasons, several researchers have performed noise reduction in the projection domain using the blind spot strategy [27,28]. These approaches satisfy the assumptions of the blind spot strategy. However, in general, as projection images for LDCT have extremely weak signals compared with reconstructed images, performing noise reconstruction using projection images is a challenging task. Therefore, these approaches indicate promising results, but the over-smoothing problems, such as smoothed edges or removed microstructures, remain.

In this work, we propose a self-supervised noise reduction method in LDCT that solves the mentioned problems. The proposed method can train a denoising neural network for LDCT without paired images. Like any other methods using the blind spot strategy, we make two statistical assumptions. First, the signal of the projection image is not pixel-wise independent. Second, the noise of the projection image is pixel-wise independent. We focus on reducing the photon noise of the projection image. Artifacts not caused by photon noise from the projection image (e.g., line defects in projection images, X-ray scattering, and artifacts caused by geometric misalignment) are not considered in this study.

Very recently, Self2Self [26] exhibited superior performance using the dropout-based ensemble. The dropout-based ensemble is a strategy that predicts noise-reduced images from randomly dropped images iteratively and outputs the result, which is the average of all predictions. It reduces the noise by dropping out the signal and restoring it using the signal from adjacent pixels. Motivated by work on Self2Self, we propose a dropped projection strategy (*DPS*) that randomly drops the pixels from the projection images, reconstructs the 3D reconstructed image using randomly dropped projection images, and trains a denoising neural network with the reconstructed images. A denoising neural network is trained to restore the signal from the reconstructed image, which is reconstructed by dropped projection images using the *DPS*, and allows the neural network to reduce the noise from the projection images. In this way, we overcome the problems of previous studies which have performed noise reduction in the projection domain, such as over-smoothing problems by using dropout in the projection domain and restoring it in the 3D reconstructed image. In addition, the denoising neural network using the *DPS* is trained to restore the signal from the pixel-wise independent noise in the projection images. This method allows noise reduction without emphasizing artifacts caused by photon noise.

The contributions of this work are as follows:

1. We define a self-supervised denoising scheme that solves the problems of noise reduction in the projection image domain and reconstructed image domain in LDCT;
2. We propose a neural network that can be trained on a small number of training samples and can yield promising results;

3. The proposed method exhibits solid performance improvement compared with the existing self-supervised noise reduction methods.

This paper is organized as follows. In Section 2, we discuss related works. In Section 3, we present a detailed description of the method and demonstrate its theoretical properties. Section 4 contains the experimental results. Finally, our summary and conclusions are provided in Section 5.

2. Related Works

Denosing is one of the essential preprocessing techniques in digital image processing. There are numerous studies on deep learning-based denoising methods for a variety of applications [29]. In particular, the denoising technique plays an important role in the medical imaging field, following the as low as reasonably achievable (ALARA) principle. Therefore, numerous studies on noise reduction in LDCT using conventional image processing techniques have been proposed to obtain good quality reconstructed images. The studies are categorized into three groups: noise reduction in the sinogram domain [3–6], iterative reconstruction [7–13], and noise reduction in the image domain [14–19].

The first group of noise reduction in LDCT is noise reduction in the sinogram domain. The key to noise reduction in the sinogram domain is filtering the noise before reconstruction. Nonlinear smoothing [3,4], bilateral filtering [5], and other statistical approaches [6] are employed for noise reduction in the sinogram domain. These methods are convenient and efficient for removing artifacts (e.g., streak artifacts). However, these methods have problems, such as a removed microstructure or blurred edges.

The second group of noise reduction in LDCT is iterative reconstruction methods. Numerous researchers have used iterative reconstruction methods for noise reduction. These methods iteratively optimize the objective function, which is defined by the tomographic system. Additionally, an iterative reconstruction method combined with a smoothing method has been proposed. Specifically, total variation [10,11], nonlocal mean filtering [11,12], dictionary learning [13], and other methods are used to improve the quality of the reconstructed image. Iterative reconstruction methods are very promising for noise reduction in LDCT, but these methods are time-consuming and hardware-intensive because of the high computational complexity. Thus, these methods are not commonly used in clinical practice.

The last group of noise reduction in LDCT is noise reduction in the image domain. The key to noise reduction in the image domain is filtering the noise after reconstruction. Total variation [14] as well as block matching and 3D filtering [15,16] are employed to reduce noise in LDCT. These methods reduce the noise efficiently but have problems, such as a removed microstructure or blurred edges, as the noise reduction in the sinogram indicated.

As an extension of these methods, methods based on deep neural networks have been proposed. The most common methods of noise reduction using deep learning are training a network with paired LDCT and normal-dose CT (NDCT) images [17–19]. The residual encoder-decoder convolutional neural network [17], wavelet residual network [18], generative adversarial network (GAN) [19], and other various network architectures are employed to reduce noise in LDCT. These methods have exhibited superior performance in reducing noise in LDCT. However, the problem with these methods is that they require numerous paired LDCT and NDCT images. Multiple scans are required to obtain well-paired LDCT and NDCT images. Multiple human body scans require additional radiation exposure. Therefore, it is almost impossible to obtain numerous well-paired LDCT and NDCT images because of public health care issues.

As research on the GAN actively progresses, methods [20–22] using a neural network trained on unpaired images have been proposed to solve the problem. These methods learn the distribution of NDCT images to reduce the noise in LDCT images. However, although the network is trained on unpaired LDCT and NDCT images, obtaining many LDCT and NDCT images is an expensive and sometimes unavailable option. For these reasons, we

propose a deep learning-based method that can train a denoising neural network on only LDCT images.

3. Methodology

This section presents a detailed description of the proposed *DPS* and discusses the self-supervised training scheme and denoising scheme.

3.1. Preliminary

Training a denoising neural network in a classical supervised manner can be performed by solving the following:

$$\operatorname{argmin}_{\theta} E \left[\| f_{\theta}(x) - y \|_2^2 \right], \tag{1}$$

where f_{θ} is a denoising neural network parameterized by θ , x is the noisy image used as an input to the neural network, and y is the clean image with the same information as the noisy image x . During training, f_{θ} is optimized to map a noisy image to a clean image and is applied in self-supervised learning-based denoising methods in the same manner. As mentioned, self-supervised learning-based denoising methods can train a neural network without paired images. These methods are commonly based on the blind spot strategy, which has exhibited promising results in the natural image domain. However, it is difficult to apply the blind spot strategy to LDCT denoising due to the assumption that the noise is pixel-wise independent. In the rest of this subsection, we describe why the noise in the LDCT reconstructed image is not pixel-wise independent.

Filtered back projection (FBP) is a widely used reconstruction algorithm in which a point of a 3D reconstructed image is reconstructed by integrating the points of filtered projections along the ray. For the convenience of explanation, in 2D parallel beam reconstruction, let us assume an object distribution function $f(x, y)$. All of the Cartesian space we used for explanation in this subsection is illustrated in Figure 2.

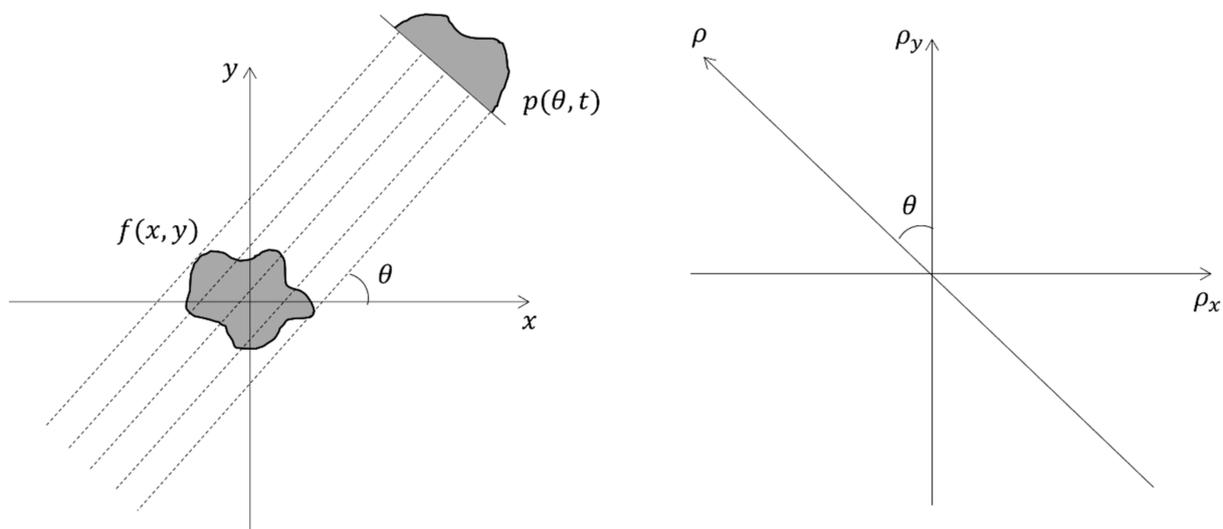


Figure 2. Tomographic system assumed for descriptive purposes. (Left) Illustration for the Radon transform theorem. (Right) Illustration for the Fourier slice theorem.

In the assumed tomographic system, a projection p can be achieved by a Radon transformation [30]. The formulation is below:

$$p(\theta, t) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f(x, y) \delta(y \cos \theta - x \sin \theta - t) dx dy, \tag{2}$$

where $\delta(\cdot)$ is Dirac’s delta function, θ is the projection angle, and t is the distance between the detector and the center of the rotation. The projection p is the ray sum along the line

through the object. The noise in the reconstructed LDCT images is caused by the photon noise in the projection images [2]. Therefore, the formula considering the noise distribution function n can be formulated as follows:

$$\hat{p}(\theta, t) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f(x, y) \delta(y \cos \theta - x \sin \theta - t) dx dy + n(\theta, t), \tag{3}$$

where \hat{p} is a parallel projection with photon noise and n is a random photon noise distribution function observed in the projection images. The projection with noise \hat{p} using the Fourier slice theorem is also defined as follows:

$$F_1 \hat{p} = F_2 \hat{f}(-\rho \sin \theta, \rho \cos \theta), \tag{4}$$

where F_n represents the n -dimensional Fourier transform and \hat{f} denotes the object distribution function reconstructed from the projections with the noise. Equation (4) describes a 1D Fourier transform of parallel projection equal to the radial line parallel to the projection in the 2D Fourier transform of the object. The object distribution function \hat{f} in the Fourier domain to obtain the reconstructed image from the projection \hat{p} is defined as follows:

$$\hat{f}(x, y) = F_2^{-1} F_2 \hat{f}(x, y) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} F_2 \hat{f}(\rho_x, \rho_y) e^{i2\pi(x\rho_x + y\rho_y)} d\rho_x d\rho_y. \tag{5}$$

Equation (5) is achieved by the Fourier inversion theorem. By inserting Equation (4) into Equation (5) and changing the variables $d\rho_x$ and $d\rho_y$ to the variable $|\rho|d\rho d\theta$, Equation (5) can be rewritten as follows (the theoretical details are proven in [31]):

$$\hat{f}(x, y) = \int_0^{2\pi} (\hat{p} * g_{\infty})(\theta, y \cos \theta - x \sin \theta) d\theta, \tag{6}$$

where g_{∞} is the ramp filter. Based on Equation (3), \hat{p} comprises the signal of projection p and the noise of projection n . Therefore, Equation (6) can be rewritten as follows:

$$\begin{aligned} \hat{f}(x, y) &= \int_0^{2\pi} ((p + n) * g_{\infty})(\theta, y \cos \theta - x \sin \theta) d\theta \\ &= \int_0^{2\pi} (p * g_{\infty})(\theta, y \cos \theta - x \sin \theta) d\theta + \int_0^{2\pi} (n * g_{\infty})(\theta, y \cos \theta - x \sin \theta) d\theta \end{aligned} \tag{7}$$

From the above, we assume that the projection data comprise a continuous signal. However, in a practical situation, the projection data are the sampled signal. The noisy projection $\hat{q}[i, k]$ sampled from the continuous projection $\hat{p}(\theta, t)$ can be defined as follows:

$$\hat{q}[i, k] = \hat{p}(\theta_i, t_k), \tag{8}$$

where

$$\theta_i = i\Delta\theta, \Delta\theta = 2\pi/N_{\theta}, i = 0, \dots, N_{\theta} - 1, \tag{9}$$

In addition, the following is true:

$$t_k = (k + 0.5)\Delta t - t_{max}, \Delta t = 2\pi/N_t, k = 0, \dots, N_t - 1, \tag{10}$$

where Δt indicates the pixel pitch of the detector and t_{max} denotes the half-length of the detector. Additionally, N_{θ} means the number of angles sampled at intervals of $\Delta\theta$, and N_t means the number of pixels sampled at intervals of Δt . Thus, Equation (6) can be rewritten as follows using the noisy sampled projection \hat{q} :

$$\begin{aligned} \hat{f}(x, y) &= \frac{2\pi}{N_{\theta}} \sum_{i=0}^{N_{\theta}-1} (\hat{q} * g_{\infty})[i, y \cos \theta - x \sin \theta] \\ &= \frac{2\pi}{N_{\theta}} \sum_{i=0}^{N_{\theta}-1} (q * g_{\infty})[i, y \cos \theta - x \sin \theta] + \frac{2\pi}{N_{\theta}} \sum_{i=0}^{N_{\theta}-1} (n_d * g_{\infty})[i, y \cos \theta - x \sin \theta], \end{aligned} \tag{11}$$

where n_d denotes the noise of the sampled projection \hat{q} . Equations (7) and (11) indicate that the ramp-filtered projection and ramp-filtered noise within the projection are integrated into the reconstructed space along the ray.

The results demonstrate why the noise in the LDCT reconstructed image is not pixel-wise independent. Therefore, the results of blind spot strategy-based noise reduction have a problem with emphasizing artifacts caused by photon noise.

One method to solve this problem is noise reduction in the projection domain. Noise reduction methods based on the blind spot strategy satisfy the assumptions of the blind spot strategy, because noise in the projection domain is pixel-wise independent. These methods reconstruct the 3D image from projection images that are already noise-reduced. However, projection images for LDCT have extremely weak signals. Thus, it is quite challenging to perform noise reduction in the projection image domain without over-smoothing. To solve these problems, we propose a *DPS*-based noise reduction method. The following subsection describes the details regarding the proposed *DPS*.

3.2. Dropped Projection Strategy

In the natural image domain, the blind spot strategy reduces noise by restoring the signal from the Bernoulli-sampled instances of the input image. The blind spot strategy defines the pixels in the input image as follows:

$$x_i = s_i + n_i, \tag{12}$$

where s_i denotes the signal of the i th pixel that is pixel-wise dependent and n_i denotes the noise of the i th pixel that is pixel-wise independent. The blind spot strategy uses Bernoulli sampling to remove the x_i , which is the i th pixel of the input image, and restores s_i from the adjacent pixels. A denoising neural network is trained using the blind spot strategy by solving

$$\operatorname{argmin}_{\theta} \sum_i L(f(x'_i, \theta), x_i), \tag{13}$$

where x'_i is the Bernoulli-sampled instance of the input image and $L(\cdot)$ denotes the L2 loss. This approach has demonstrated promising results in the natural image domain. However, for the mentioned reason, another approach is needed to reduce the noise of the LDCT reconstructed images. Therefore, we propose a strategy to generate the Bernoulli-sampled instance of the projection image and restore it in the reconstructed image domain. The diagram of the proposed method is presented in Figure 3.

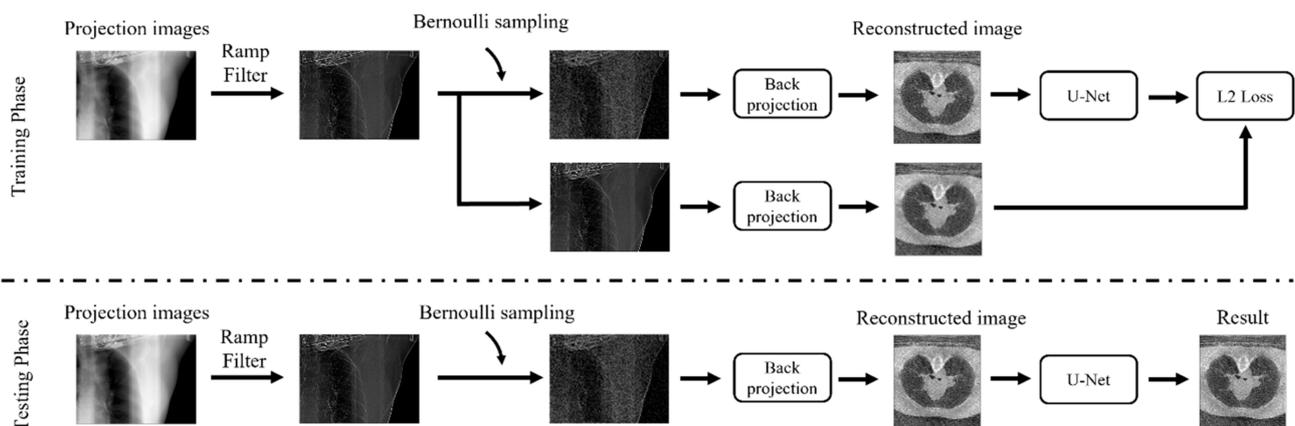


Figure 3. Flowchart of the proposed method with the dropped projection strategy (*DPS*). In the training phase, we map each slice of the reconstructed image (reconstructed using dropped projection images) to each slice of the reconstructed image (reconstructed by non-dropped projection images).

To reduce the noise in the reconstructed image domain using the blind spot strategy, we must generate the input images of the denoising neural network in which the independent

noise is removed. According to Equation (11), photon noise in the projection image is ramp-filtered and integrated into the reconstructed image along the ray. Thus, we must remove the pixels of the projection image to generate the projection image in which the independent noise is removed using Bernoulli sampling. The Bernoulli-sampled projection image q' can be defined as follows:

$$q' = (\hat{q} * g_\infty) \odot b, \tag{14}$$

where \odot represents the element-wise multiplication and b denotes the binary Bernoulli vector with elements 1 and 0, which are randomly generated. The binary Bernoulli vector b can be expressed as follows:

$$b = \begin{cases} 1 & \text{with probability } k, \\ 0 & \text{with probability } 1 - k, \end{cases} \tag{15}$$

where the probability k has a value from 0 to 1. Therefore, we can generate the 3D image f' reconstructed from the Bernoulli-sampled projection image q' . Using Equations (11) and (14), f' is defined as follows:

$$\begin{aligned} f'(x, y) &= \frac{2\pi}{N_\theta} \sum_{i=0}^{N_\theta-1} ((\hat{q} * g_\infty) \odot b)[i, y\cos\theta - x\sin\theta] \\ &= \frac{2\pi}{N_\theta} \sum_{i=0}^{N_\theta-1} q'[i, y\cos\theta - x\sin\theta]. \end{aligned} \tag{16}$$

We can generate blind spots in the reconstructed image f' using this method.

3.3. Training Scheme

In the previous subsection, we described how to make the reconstructed image from the Bernoulli-sampled projection image and the theoretical properties of the proposed method. In this subsection, we describe the scheme for training a denoising neural network using this method. To train a denoising neural network only on the images from LDCT scans, we generated a set of binary vectors using binary Bernoulli sampling. Depending on the probability k , the distribution of the randomly generated binary Bernoulli vector b has a value of 0 or 1.

In the training scheme, the denoising neural network is trained to restore the residual part of the image reconstructed without Bernoulli sampling on the projection image from the image reconstructed using Bernoulli sampling on the projection image. The residual part of the reconstructed image \bar{f} can be expressed as follows:

$$\bar{f}(x, y) = \frac{2\pi}{N_\theta} \sum_{i=0}^{N_\theta-1} ((\hat{q} * g_\infty) \odot (1 - b))[i, y\cos\theta - x\sin\theta]. \tag{17}$$

Using these methods to obtain the necessary image pairs $\{\hat{f}_m, \bar{f}_m\}_{m=0}^M$ for learning, the denoising neural network is learned to minimize the following loss function:

$$L = \sum_{m=0}^{M-1} \sum_{s=0}^{S-1} \| F_\theta(f'_{m,s}) - \bar{f}_{m,s} \|^2, \tag{18}$$

where F_θ is a denoising neural network parameterized by θ and s is the slice index of the reconstructed image. The denoising neural network trained by the proposed method can avoid convergence to an identity mapping. In addition, due to the randomly generated binary Bernoulli vector, the overfitting problem can be avoided even if the denoising neural network is trained on a small number of LDCT scans.

3.4. Denoising Scheme

The proposed method reconstructs the 3D image using Bernoulli-sampled projection and trains the denoising neural network to restore the signal removed by Bernoulli sam-

pling. In the proposed method, the pixel signals are randomly removed by the Bernoulli sampling probability k . Therefore, the denoising neural network does not reduce the noise of all pixels at once. To solve this problem, we generate multiple outputs of the neural network with independent binary Bernoulli vectors and use the average of multiple outputs as the final results. The final results f^* can be defined with the average of the multiple outputs f'_0, f'_1, \dots, f'_n generated using the independent Bernoulli vector b_0, b_1, \dots, b_n as follows:

$$f^* = \frac{1}{N} \sum_{n=0}^{N-1} f'_n. \quad (19)$$

4. Experiments and Results

In this section, we evaluate the performance of the proposed method. We also introduce the implementation details of the proposed method.

The proposed method was evaluated on two datasets: (1) the SPARE challenge dataset [32], which is the clinical sparse-view 4D CBCT dataset, and (2) the lithium polymer battery dataset with the in-house CBCT system. Both datasets provide projection images and the geometry of the tomographic system. The proposed method had promising results in both the SPARE challenge dataset and lithium polymer battery dataset.

4.1. Evaluation Metrics

To measure the performance of the proposed method and other methods for comparison, we used the peak signal-to-noise ratio (PSNR) and structural similarity index measure (SSIM) [33] as evaluation metrics.

4.1.1. PSNR

To measure the similarity of the pixel values between the denoised reconstructed images and the target images, we used PSNR, which was defined as

$$PSNR = 10 \cdot \log \frac{s^2}{MSE}, \quad (20)$$

where s means the maximum value of the target image and MSE is defined as

$$MSE = \| f_\theta(x) - y \|_2^2. \quad (21)$$

4.1.2. SSIM

To measure the structural similarity of the reconstructed images, we used SSIM, which was defined as

$$SSIM = \frac{2\mu_{f_\theta(x)}\mu_y + c_1}{\mu_{f_\theta(x)}^2 + \mu_y^2 + c_1} \cdot \frac{2\sigma_{f_\theta(x)y} + c_2}{\sigma_{f_\theta(x)}^2 + \sigma_y^2 + c_2}, \quad (22)$$

where $\mu_{f_\theta(x)}$ and μ_y are the means, $\sigma_{f_\theta(x)}$ and σ_y are the standard deviations of the reconstructed image $f_\theta(x)$ and y , respectively, $\sigma_{f_\theta(x)y}$ is the covariance of the reconstructed image $f_\theta(x)$ and y , and

$$\begin{aligned} c_1 &= (k_1 L)^2, \\ c_2 &= (k_2 L)^2. \end{aligned} \quad (23)$$

where L means the maximum value of the target image and k_1 and k_2 are 0.01 and 0.03, respectively.

4.2. Implementation Details

In an experiment to evaluate the performance of the proposed methods, we used the "U-Net"-based architecture proposed in Pix2Pix [34]. The differences between the original U-Net architecture [35] and the U-Net-based architecture are that several skip connections between layers were added, batch normalization was used, and dropout layers were added to the decoding part. We used the Adam [36] optimizer with a linearly decaying learning

rate from 1×10^{-4} to 0. The dropout probability of all dropout layers was set to 0.3, and the Bernoulli sampling probability k was set to 0.3.

To improve the performance of the neural network, we also used data augmentation, including horizontal flipping, vertical flipping, and diagonal flipping, and we used $N = 50$ in Equation (19) for the denoising scheme. To evaluate the performance of the proposed method, we used recent deep learning-based self-supervised denoising methods and conventional denoising methods as counterpart methods: $N2V$ [24], $N2S$ [25], nonlocal mean filtering, total variation, and bilateral filtering. These methods were applied in the projection image domain. All of the counterpart denoising neural networks were trained using published training code.

4.3. SPARE Challenge Dataset

The SPARE challenge dataset is a clinical sparse-view 4D CBCT dataset. We evaluated the proposed method and counterpart methods on the CV_P1_T_01 dataset, a subset of the SPARE challenge dataset. To evaluate the performance of reducing the noise in the LDCT scan, we regarded the CT scan with half of the projections as LDCT and the CT scan with all of the projections as NDCT. Using half of the projections to reconstruct the 3D image indicates that half of the dose is irradiated.

The CV_P1_T_01 dataset is a 4D CT dataset. Thus, it can be divided into 10 subdatasets, depending on the respiratory phase. Therefore, all image data in the SPARE dataset have a respiratory phase value from 1 to 10. We used image data with a respiratory phase value of 1 as the testing data, and the rest of the data with respiratory phase values of 2–9 were used as training data. We built the proposed models in two types: DPS^s and DPS^d . DPS^s was trained using half of the projection data with a respiratory phase value of one and was tested on the same dataset used in the training phase to demonstrate that it could be trained efficiently on a small dataset, like single-image denoising, as the usual practice in self-supervised denoising methods [24,26]. DPS^d and other deep learning-based methods ($N2V$ and $N2S$) were trained using 2039 projection images with a size of 1024×768 (images with response phase values of 2–9).

The method for evaluating the performance of each approach was as follows. For the counterpart methods, the 3D data used for the performance evaluation were reconstructed by the FBP-based reconstruction algorithm with half of the denoised projection images, with a respiratory phase value of one. For the proposed method, the 3D data used for the performance evaluation were generated by the method described in Section 3 with half of the projection images, with a respiratory phase value of one. The 3D images generated in these ways were compared to the 3D images reconstructed with all projection images, with a respiratory phase value of one. The size of the reconstructed 3D images was $448 \times 160 \times 448$.

Table 1 and Figure 4 present the performance comparison. Table 1 lists the quantitative results of various denoising methods, including the proposed method. In Table 1, single image set learning or non-learning methods means the deep learning-based denoising methods that were trained with a single image set or conventional image processing-based denoising methods. Dataset-based deep learning method refers to the deep learning-based denoising methods that were trained with multiple sets of images.

Table 1. Quantitative results from the SPARE challenge dataset.

	Single Image Set Learning or Non-Learning Methods			Dataset-Based Deep Learning Method		
	<i>NLM</i>	<i>TV</i>	<i>DPS^s</i>	<i>N2S</i>	<i>N2V</i>	<i>DPS^d</i>
PSNR (dB)	24.06	26.21	30.680	23.97	25.83	30.667
SSIM	0.55	0.65	0.787	0.56	0.67	0.786

NLM = non-local mean filtering; *TV* = total variation; *N2S* = Noise2Self [25]; *N2V* = Noise2Void [26]. *DPS^s* and *DPS^d* are the proposed methods. *DPS^s* is trained with a single image set, and *DPS^d* is trained with multiple sets of images.

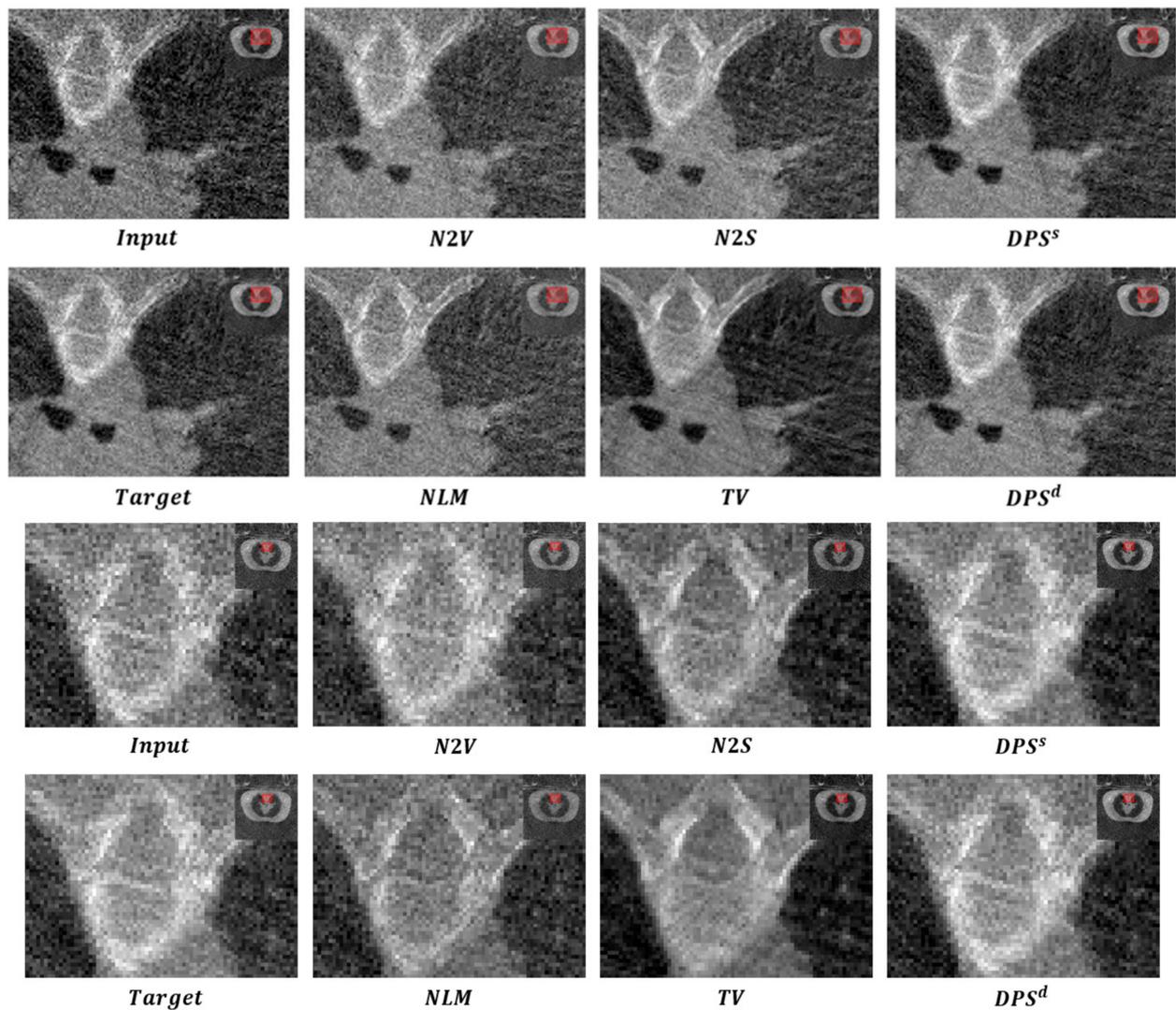


Figure 4. Two visual comparisons of various denoising methods from SPARE challenge dataset: (first and second rows) coarse architecture and (third and fourth rows) fine architecture. All images are in the axial view. Red boxes in the images indicate the expanded regions.

The method performed better than the counterparts (i.e., non-learning- and learning-based methods). Moreover, the proposed method could train a denoising neural network efficiently, even on a small dataset, as DPS^s had comparable results to DPS^d . Figure 4 depicts the visual comparison of various denoising methods. The microstructures were blurred, and some artifacts occurred in the results of the counterpart methods. However, the results of the proposed method exhibited very comparable quality with the target images. Therefore, the proposed method had promising results both visually and quantitatively.

4.4. Lithium Polymer Battery Dataset

We also evaluated our proposed method and the counterpart methods on the lithium polymer battery dataset with an in-house CBCT system. In the same manner as the experiment on the SPARE dataset, we evaluated the proposed method and counterpart methods.

However, in the experiments on the lithium polymer battery dataset, we trained the neural networks on 200 projection images of 1248×448 in the proposed method (DPS) and other deep learning-based methods ($N2V$ and $N2S$) and tested them on the same dataset which was used in the training phase. For each denoising method, the 3D images reconstructed with 200 denoised projection images were compared to the 3D image

reconstructed with 400 projection images. The size of the reconstructed 3D images was $384 \times 384 \times 400$.

Table 2 and Figure 5 present the performance comparison. Table 2 lists the quantitative results of various denoising methods, including the proposed method. In the quantitative results, the counterpart deep learning-based methods performed poorly, as unknown artifacts were generated in the reconstructed 3D image (see Figure 6). The projection images for CT had extremely weak signals. Therefore, unknown artifacts could be generated during denoising, and these artifacts were emphasized in the reconstructed image.

Table 2. Quantitative results from the lithium polymer dataset.

	Single Image Set Learning or Non-Learning Methods					
	<i>Bilateral</i>	<i>NLM</i>	<i>TV</i>	<i>N2V</i>	<i>N2S</i>	<i>DPS</i>
PSNR(dB)	17.44	23.27	24.01	21.15	20.31	30.68
SSIM	0.21	0.70	0.64	0.37	0.24	0.79

Bilateral = bilateral filtering.

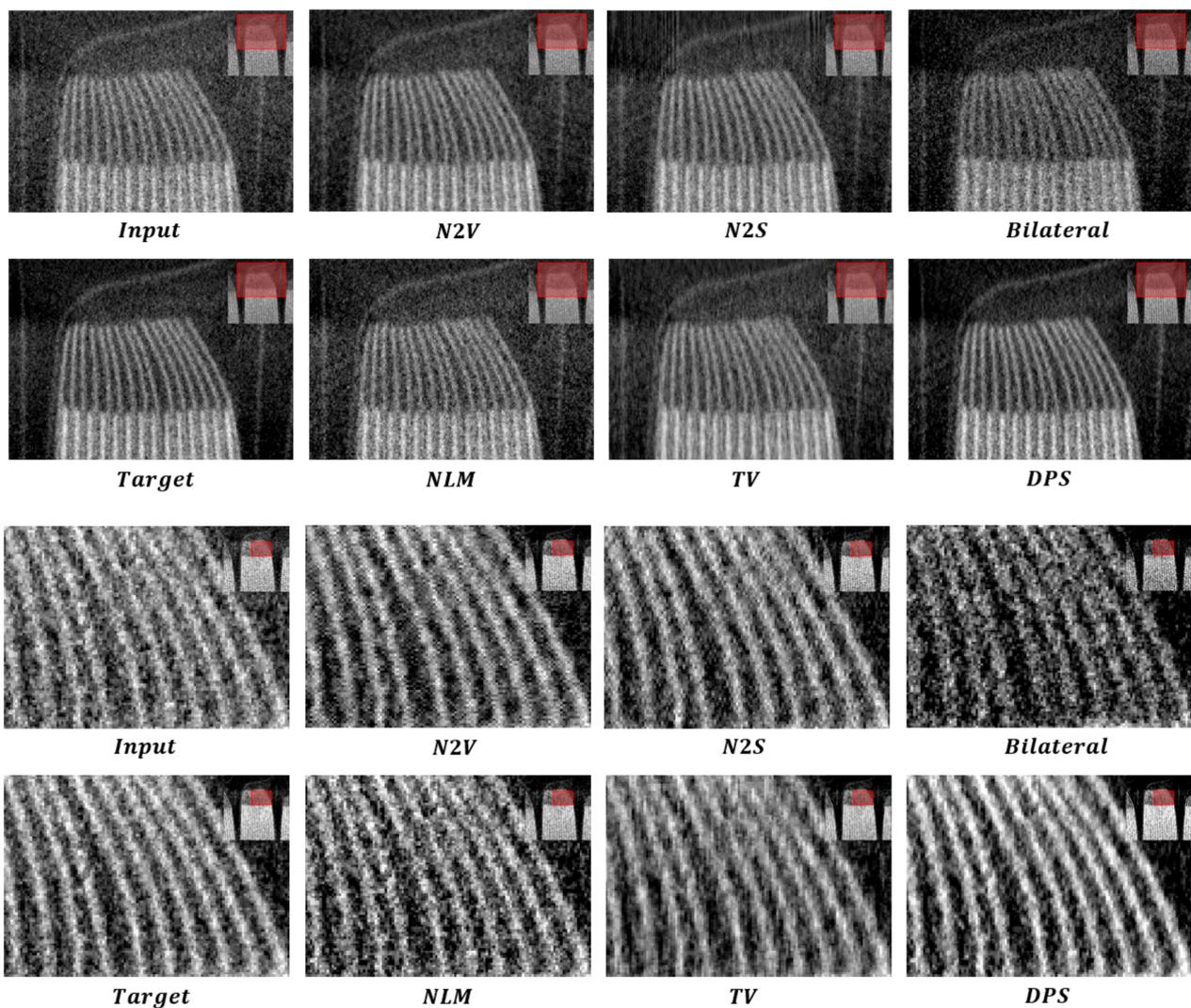


Figure 5. Two visual comparisons of various denoising methods from the lithium polymer battery dataset with an in-house CBCT system: (first and second rows) coarse architecture and (third and fourth rows) fine architecture. All images are in the coronal view. Red boxes in the images indicate the expanded regions.

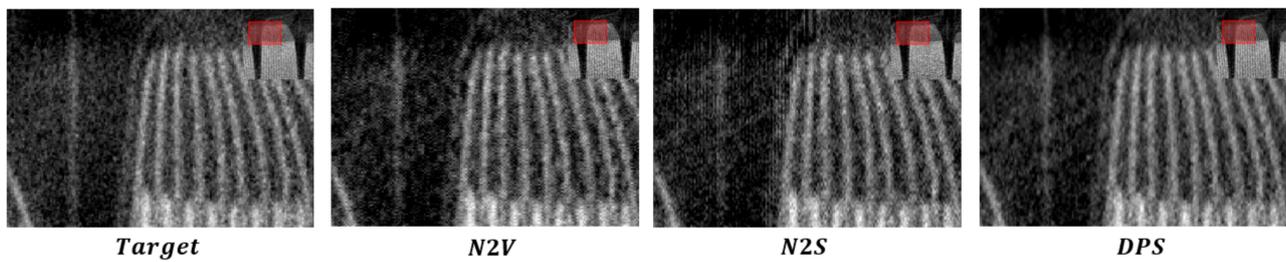


Figure 6. Artifacts in the results of the counterpart deep learning-based methods. All images are in the coronal view. Red boxes in the images indicate the expanded regions.

However, the proposed method applied denoising on the reconstructed image domain, and the artifacts did not appear. Thus, the proposed method exhibited promising performance both quantitatively and visually. Figure 5 depicts the visual comparison of various denoising methods. The image quality was comparable to the target image. The distinction between anodes was especially better in the proposed method than the counterpart methods.

4.5. Performance for Repetition

In the denoising scheme subsection, we described the average of the multiple outputs used as the final outputs of the proposed method with Equation (19). In this subsection, we analyze the effects of the number of outputs on the performance of the method. Figure 7 displays the changes in the PSNR and SSIM values according to the number of outputs. The PSNR and SSIM exhibited similar trends according to the number of outputs and started to converge at about $N = 25$ in Equation (19). By adjusting the number of iterations, we could practically obtain the result within an acceptable processing time.

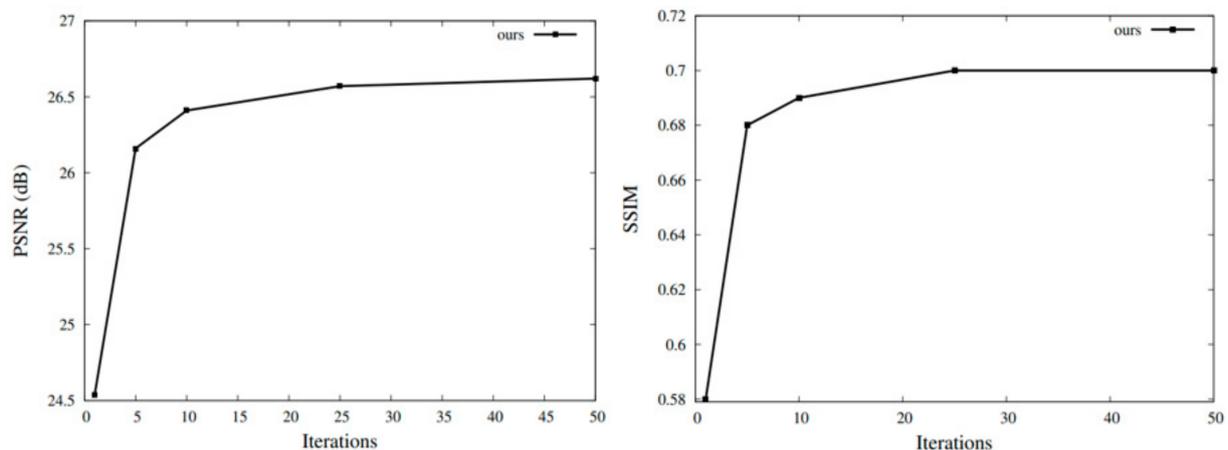


Figure 7. Peak signal-to-noise ratio (PSNR) and structural similarity index measure (SSIM) versus the number of iterations. Both the PSNR and SSIM increased with the number of iterations and gradually converged.

5. Conclusions

We proposed a self-supervised learning-based method that can train a denoising neural network for LDCT without paired images. The proposed method uses Bernoulli sampling to generate degraded versions of the projection images and reconstruct the 3D image, and the denoising neural network is trained to restore the image dropped out by Bernoulli sampling in the projection image domain. The proposed method can mitigate the over-smoothing problem in conventional methods with small data. To verify the performance of the proposed method, we compared its performance with various counterpart methods, including deep learning-based methods ($N2V$, $N2S$, and $S2S$) and non-learning methods

(bilateral filtering, *TV*, and *NLM*). Quantitatively, the results of our experiments showed that our proposed method outperformed the counterpart denoising methods by at least 4.47 dB in terms of PSNR value on the SPARE challenge dataset and the in-house lithium polymer dataset. In particular, we solved the problem of generating artifacts and the over-smoothing when applying the denoising method in the projection domain. This method can be applied to various tomographic systems in medical and industrial areas.

Author Contributions: Conceptualization, Y.-J.H.; methodology, Y.-J.H.; software, Y.-J.H.; validation, Y.-J.H.; formal analysis, Y.-J.H.; investigation, Y.-J.H.; resources, Y.-J.H.; data curation, Y.-J.H.; writing—original draft preparation, Y.-J.H.; writing—review and editing, H.-J.Y.; visualization, Y.-J.H.; supervision, H.-J.Y.; project administration, H.-J.Y.; funding acquisition, H.-J.Y. All authors have read and agreed to the published version of the manuscript.

Funding: This research was supported by the Basic Science Research Program through the National Research Foundation of Korea (NRF), funded by the Ministry of Science, ICT, and Future Planning (2020R1A2C1007081).

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Conflicts of Interest: The authors declare no conflict of interest.

References

- Brenner, D.J.; Hall, E.J. Cancer risks from CT scans: Now we have data, what next? *Radiology* **2012**, *265*, 330–331. [[CrossRef](#)] [[PubMed](#)]
- Lee, S.; Lee, M.S.; Kang, M.G. Poisson–Gaussian noise analysis and estimation for low-dose X-ray images in the NSCT domain. *Sensors* **2018**, *18*, 1019. [[CrossRef](#)] [[PubMed](#)]
- Wang, J.; Lu, H.; Li, T.; Liang, Z. Sinogram noise reduction for low-dose CT by statistics-based nonlinear filters. In Proceedings of the Medical Imaging 2005, San Diego, CA, USA, 13–17 February 2005; pp. 2058–2066.
- Li, T.; Li, X.; Wang, J.; Wen, J.; Lu, H.; Hsieh, J.; Liang, Z. Nonlinear sinogram smoothing for low-dose X-ray CT. *IEEE Trans. Nucl. Sci.* **2004**, *51*, 2505–2513.
- Yu, L.; Manduca, A.; Trzasko, J.D.; Khaylova, N.; Kofler, J.M.; McCollough, C.M.; Fletcher, J.G. Sinogram smoothing with bilateral filtering for low-dose CT. In Proceedings of the Medical Imaging 2008, San Diego, CA, USA, 17–19 February 2008; pp. 768–775.
- La Rivière, P.J.; Bian, J.; Vargas, P.A. Penalized-likelihood sinogram restoration for computed tomography. *IEEE Trans. Med. Imaging* **2006**, *25*, 1022–1036. [[CrossRef](#)] [[PubMed](#)]
- Elbakri, I.A.; Fessler, J.A. Fessler. Statistical image reconstruction for polyenergetic X-ray computed tomography. *IEEE Trans. Med. Imaging* **2002**, *21*, 89–99. [[CrossRef](#)] [[PubMed](#)]
- Wang, J.; Li, T.; Xing, L. Iterative image reconstruction for CBCT using edge-preserving prior. *Med. Phys.* **2008**, *36*, 252–260. [[CrossRef](#)]
- Beister, M.; Kolditz, D.; Kalender, W.A. Iterative reconstruction methods in X-ray CT. *Phys. Med.* **2012**, *28*, 94–108. [[CrossRef](#)]
- Luo, X.; Yu, W.; Wang, C. An image reconstruction method based on total variation and wavelet tight frame for limited-angle CT. *IEEE Access* **2017**, *6*, 1461–1470. [[CrossRef](#)]
- Ertas, M.; Yildirim, I.; Kamasak, M.; Akan, A. An iterative tomosynthesis reconstruction using total variation combined with non-local means filtering. *Biomed. Eng. Online* **2014**, *13*, 65. [[CrossRef](#)]
- Kelm, Z.S.; Blezek, D.; Bartholmai, B.; Erickson, B.J. Optimizing non-local means for denoising low dose CT. In Proceedings of the 2009 IEEE International Symposium on Biomedical Imaging: From Nano to Macro, Boston, MA, USA, 28 June–1 July 2009; pp. 662–665.
- Xu, Q.; Yu, H.; Mou, X.; Zhang, L.; Hsieh, J.; Wang, G. Low-dose X-ray CT reconstruction via dictionary learning. *IEEE Trans. Med. Imaging* **2012**, *31*, 1682–1697.
- Li, Z.; Yu, L.; Trzasko, J.D.; Fletcher, J.G.; McCollough, C.H.; Manduca, A. Adaptive non-local means filtering based on local noise level for CT denoising. In Proceedings of the Medical Imaging 2012, San Diego, CA, USA, 5–7 February 2012; pp. 447–456.
- Kang, D.; Slomka, P.; Nakazato, R.; Woo, J.; Berman, D.S.; Kuo, C.-C.J.; Dey, D. Image denoising of low-radiation dose coronary CT angiography by an adaptive block-matching 3D algorithm. In Proceedings of the Medical Imaging 2013, Lake Buena Vista, FL, USA, 10–11 February 2013; pp. 671–676.
- Hasan, A.M.; Melli, A.; Wahid, K.A.; Babyn, P. Denoising low-dose CT images using multiframe blind source separation and block matching filter. *IEEE Trans. Radiat. Plasma Med. Sci.* **2018**, *2*, 279–287. [[CrossRef](#)]
- Chen, H.; Zhang, Y.; Kalra, M.K.; Lin, F.; Chen, Y.; Liao, P.; Zhou, J.; Wang, G. Low-dose CT with a residual encoder-decoder convolutional neural network. *IEEE Trans. Med. Imaging* **2017**, *36*, 2524–2535. [[CrossRef](#)] [[PubMed](#)]

18. Kang, E.; Chang, W.; Yoo, J.; Ye, J.C. Deep convolutional framelet denoising for low-dose CT via wavelet residual network. *IEEE Trans. Med. Imaging* **2018**, *37*, 1358–1369. [[CrossRef](#)] [[PubMed](#)]
19. Yang, Q.; Yan, P.; Zhang, Y.; Yu, H.; Shi, Y.; Mou, X.; Kalra, M.K.; Zhang, Y.; Sun, L.; Wang, G. Low-dose CT image denoising using a generative adversarial network with Wasserstein distance and perceptual loss. *IEEE Trans. Med. Imaging* **2018**, *37*, 1348–1357. [[CrossRef](#)]
20. Tang, C.; Li, J.; Wang, L.; Li, Z.; Jiang, L.; Cai, A.; Zhang, W.; Liang, N.; Li, L.; Yan, B. Unpaired low-dose CT denoising network based on cycle-consistent generative adversarial network with prior image information. *Comput. Math. Methods Med.* **2019**, *2019*, 8639825. [[CrossRef](#)]
21. Park, H.S.; Baek, J.; You, S.K.; Choi, J.K.; Seo, J.K. Unpaired image denoising using a generative adversarial network in X-ray CT. *IEEE Access* **2019**, *7*, 110414–110425. [[CrossRef](#)]
22. Li, Z.; Zhou, S.; Huang, J.; Yu, L.; Jin, M. Investigation of low-dose CT image denoising using unpaired deep learning methods. *IEEE Trans. Radiat. Plasma Med. Sci.* **2020**, *5*, 224–234. [[CrossRef](#)]
23. Lehtinen, J.; Munkberg, J.; Hasselgren, J.; Laine, S.; Karras, T.; Aittala, M.; Aila, T. Noise2noise: Learning image restoration without clean data. *arXiv preprint* **2018**, arXiv:1803.04189.
24. Krull, A.; Buchholz, T.O.; Jug, F. Noise2void-learning denoising from single noisy images. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Long Beach, CA, USA, 15–20 June 2019; pp. 2129–2137.
25. Batson, J.; Royer, L. Noise2self: Blind denoising by self-supervision. In Proceedings of the International Conference on Machine Learning, Long Beach, CA, USA, 10–15 June 2019; pp. 524–533.
26. Quan, Y.; Chen, M.; Pang, T.; Ji, H. Self2self with dropout: Learning self-supervised denoising from single image. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Online, 14–19 June 2020; pp. 1890–1898.
27. Liang, K.; Zhang, L.; Xing, Y. Training a low-dose CT denoising network with only low-dose CT dataset: Comparison of DDLN and Noise2Void. In Proceedings of the Medical Imaging 2021, Online, 15–20 February 2021; p. 1159501.
28. Unal, M.O.; Ertas, M.; Yildirim, I. Self-Supervised Training for Low-Dose Ct Reconstruction. In Proceedings of the 2021 IEEE 18th International Symposium on Biomedical Imaging (ISBI), Online, 13–16 April 2021; pp. 69–72.
29. Tian, C.; Fei, L.; Zheng, W.; Xu, Y.; Zuo, W.; Lin, C.-W. Deep learning on image denoising: An overview. *Neural Netw.* **2020**, *131*, 251–275. [[CrossRef](#)]
30. Radon, J. On the determination of functions from their integral values along certain manifolds. *IEEE Trans. Med. Imaging* **1986**, *5*, 170–176. [[CrossRef](#)]
31. Turbell, H. Cone-Beam Reconstruction Using Filtered Backprojection. Ph.D. Thesis, Linköping University, Linköping, Sweden, 2001.
32. Shieh, C.-C.; Gonzalez, Y.; Li, B.; Jia, X.; Rit, S.; Mory, C.; Riblett, M.; Hugo, G.; Zhang, Y.; Jiang, Z.; et al. SPARE: Sparse-view reconstruction challenge for 4D cone-beam CT from a 1-min scan. *Med. Phys.* **2019**, *46*, 3799–3811. [[CrossRef](#)] [[PubMed](#)]
33. Wang, Z.; Bovik, A.C.; Sheikh, H.R.; Simoncelli, E.P. Image quality assessment: From error visibility to structural similarity. *IEEE Trans. Image Process.* **2004**, *13*, 600–612. [[CrossRef](#)] [[PubMed](#)]
34. Isola, P.; Zhu, J.Y.; Zhou, T.; Efros, A.A. Image-to-image translation with conditional adversarial networks. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, HI, USA, 21–26 July 2017; pp. 1125–1134.
35. Ronneberger, O.; Fischer, P.; Brox, T. U-net: Convolutional networks for biomedical image segmentation. In Proceedings of the International Conference on Medical Image Computing and Computer-Assisted Intervention, Munich, Germany, 5–9 October 2015; pp. 234–241.
36. Kingma, D.P.; Ba, J. Adam: A method for stochastic optimization. *arXiv preprint* **2014**, arXiv:1412.6980.