

PAPER


Self-supervised dual-domain balanced dropblock-network for low-dose CT denoising

To cite this article: Ran An *et al* 2024 *Phys. Med. Biol.* **69** 075026


View the [article online](#) for updates and enhancements.

You may also like

- [An improved dual-domain network for metal artifact reduction in CT images using aggregated contextual transformations](#)
Hui Tang, Sudong Jiang, Yubing Lin *et al.*
- [Mud-Net: multi-domain deep unrolling network for simultaneous sparse-view and metal artifact reduction in computed tomography](#)
Baoshun Shi, Ke Jiang, Shaolei Zhang *et al.*
- [DAN-Net: Dual-domain adaptive-scaling non-local network for CT metal artifact reduction](#)
Tao Wang, Wenjun Xia, Yongqiang Huang *et al.*



Joining forces:
One complete
QA solution for
Dosimetry with
myQA[®], QUASAR[™]
and Radcal[®]!



The diagram is a circular graphic with a dark background and a pattern of small, colorful dots. It consists of four concentric rings. The outermost ring is divided into three segments: a dark blue segment labeled 'Machine QA', a green segment labeled 'Patient Specific QA', and a light blue segment labeled 'Medical Imaging QA'. The middle ring is a solid pink circle labeled 'Risk Management'. The center of the diagram features a white wireframe model of a human head and neck, rendered in a light green color.



PAPER

Self-supervised dual-domain balanced dropblock-network for low-dose CT denoising

RECEIVED
6 November 2023REVISED
22 January 2024ACCEPTED FOR PUBLICATION
15 February 2024PUBLISHED
26 March 2024Ran An^{1,2}, Ke Chen^{3,*} and Hongwei Li^{1,*}¹ School of Mathematical Sciences, Capital Normal University, Beijing, 100048, People's Republic of China² Centre for Mathematical Imaging Techniques, University of Liverpool, Liverpool, L69 7ZL, United Kingdom³ Department of Mathematics and Statistics, University of Strathclyde, Glasgow, G1 1XQ, United Kingdom

* Authors to whom any correspondence should be addressed.

E-mail: k.chen@strath.ac.uk and hongwei.li91@cnu.edu.cn**Keywords:** low-dose CT denoising, self-supervised learning, dual-domain denoising, sinogram denoising**Abstract**

Objective. Self-supervised learning methods have been successfully applied for low-dose computed tomography (LDCT) denoising, with the advantage of not requiring labeled data. Conventional self-supervised methods operate only in the image domain, ignoring valuable priors in the sinogram domain. Recently proposed dual-domain methods address this limitation but encounter issues with blurring artifacts in the reconstructed image due to the inhomogeneous distribution of noise levels in low-dose sinograms. *Approach.* To tackle this challenge, this paper proposes SDBDNet, an end-to-end dual-domain self-supervised method for LDCT denoising. With the network designed based on the properties of inhomogeneous noise in low-dose sinograms and the principle of moderate sinogram-domain denoising, SDBDNet achieves effective denoising in dual domains without introducing blurring artifacts. Specifically, we split the sinogram into two subsets based on the positions of detector cells to generate paired training data with high similarity and independent noise. These sub-sinograms are then restored to their original size using 1D interpolation and learning-based correction. To achieve adaptive and moderate smoothing in the sinogram domain, we integrate Dropblock, a type of convolution layer with regularization, into SDBDNet, and set a weighted average between the denoised sinograms and their noisy counterparts, leading to a well-balanced dual-domain approach. *Main results.* Numerical experiments show that our method outperforms popular non-learning and self-supervised learning methods, demonstrating its effectiveness and superior performance. *Significance.* While introducing a novel high-performance dual-domain self-supervised LDCT denoising method, this paper also emphasizes and verifies the importance of appropriate sinogram-domain denoising in dual-domain methods, which might inspire future work.

1. Introduction

Computed tomography (CT) is an important imaging technique in medical diagnosis. However, excessive x-ray doses from CT scans can potentially harm patients. To mitigate this risk, it is necessary to reduce the x-ray dose. In the field of medical CT, the 'as Low as Reasonably Achievable' (ALARA) guideline (de González *et al* 2010) was introduced to minimize radiation exposure during CT diagnosis. Nonetheless, low-dose CT (LDCT) tends to introduce noticeable noise and artifacts in the reconstructed images compared to normal-dose CT. Hence, LDCT denoising has always been a hot research topic in the medical imaging community.

Denoising the reconstructed image or sinogram directly with classical image denoising methods is convenient and practical. For instance, local averaging methods (Yaroslavsky 2012) based on smoothing filters, and nonlocal averaging methods like NLM (Buades *et al* 2005) and BM3D (Dabov *et al* 2007), which rely on image internal similarity priors, can be employed for LDCT denoising. These methods are easy to implement and fast to run based on explicit degradation models. However, the noise characteristics of both sinograms and

images of LDCT do not fully conform to the assumptions of these methods, making it challenging for them to achieve satisfactory results.

In recent years, deep learning methods with convolutional neural networks (CNNs) have been proposed for image denoising with great success. Popular network architectures include DnCNN (Zhang *et al* 2017), FFD-Net (Zhang *et al* 2018), CBD-Net (Guo *et al* 2019), and many variations of UNet (Ronneberger *et al* 2015). They achieve better results than traditional methods while using significantly less time for inference. In the field of LDCT denoising, numerous networks with different structures have been proposed. Representative examples include FBPCnnNet (Jin *et al* 2017), RED-CNN (Chen *et al* 2017) and CTformer (Wang *et al* 2023), which achieve powerful denoising effects on LDCT images. Some other methods perform supervised learning denoising in both the sinogram domain and the image domain with a reconstruction operator for domain transformation, thus achieving excellent denoising performance. Examples of the representative methods are DDPNet (Ge *et al* 2022), DuDoUFNet (Zhou *et al* 2022) and CLEAR (Zhang *et al* 2021).

Although deep learning methods have shown exemplary performance in image denoising tasks, mainstream supervised denoising methods require a large amount of labeled data for network training, which is difficult to collect in real applications. In fact, in clinical diagnosis, it is impractical to scan the same patient twice at different doses to obtain paired data. Additionally, other factors, such as uncontrollable motion due to respiration, further prevent high-quality labeled data from being collected. To obtain labeled and paired data more easily, some generative methods have been utilized in LDCT denoising. Some of them generate data by physical models. For example, Zeng *et al* simulated low-dose sinograms from high-dose sinograms using a CT projection model to obtain paired data (Zeng *et al* 2015). However, the generated data from such methods is not of high quality, and it remains a big challenge to generate realistic data of CT. Some others of the generative methods directly generate data through a deep-learning model. The most representative ones are based on GAN (Creswell *et al* 2018), such as GAN-CIRCLE (You *et al* 2019), Cycle-Free CycleGAN (Kwon and Ye 2021), AdaIN-Based Tunable CycleGAN (Gu and Ye 2021), IdentityGAN (Li *et al* 2020) and NE-GAN (Niu *et al* 2021). These GAN-based methods use unpaired noisy-clean data, avoiding the need for paired data successfully. However, most of them rely on the cyclic consistency loss of GAN, which usually causes poor stability during training. Moreover, ensuring the accuracy of these generative models in LDCT denoising is not easy. Therefore, the application of supervised denoising methods for practical diagnosis is limited.

To avoid the reliance on labeled data, many unsupervised learning methods have been proposed in recent years. These methods are more practical and have also achieved good results in image denoising. One representative method is Noise2Noise (N2N) (Lehtinen *et al* 2018), which trains networks with paired noisy images. N2N shows that training with two images of the same scene with uncorrelated and zero-mean noise can also achieve image denoising. Inspired by N2N, several methods were proposed for LDCT denoising (Hasan *et al* 2020, Won *et al* 2021, Wu *et al* 2021, Fang *et al* 2021, Yuan *et al* 2020, Zhang *et al* 2022, 2021). However, an inevitable problem when applying these methods in LDCT denoising is that paired two low-dose CT images are hard to obtain. Another class of methods learns priors in only normal-dose CT images by generative models and incorporates them into low-dose CT reconstruction. Fabian *et al* proposed PatchNR (Altekrüger *et al* 2023) to learn a regularization term from the patches of a few normal-dose images with normalizing flows (Kobyzev *et al* 2020) and used it in iterative low-dose reconstruction. He *et al* utilized a score-based model (Song and Ermon 2020) to learn priors in normal-dose images and conducted iterative reconstruction, thus proposing EASEL (He *et al* 2022). Liu *et al* proposed Dn-Dp (Liu *et al* 2023) which adopts diffusion models (Croitoru *et al* 2023) to learn priors in normal-dose images and solve multiple maximum-a-posteriori problems iteratively to achieve denoising for low-dose images. Such methods only require normal-dose images for training and achieve effective LDCT denoising. However, the requirement for a large amount of normal-dose images still limits their application to situations with only low-dose data. In addition, such methods usually require iterative solving, which often means larger time consumption.

Recently, self-supervised learning methods have also been proposed for image denoising which explore and utilize the local similarity of the noisy image itself. Deep image prior (DIP) (Ulyanov *et al* 2018) leverages the regularization effect of networks and achieves denoising by early terminating the training process of a generative network. Noise2Void (Krull *et al* 2019) and Noise2Self (Batson and Royer 2019) were proposed for image denoising by predicting pixel values based on their local noisy surroundings, under the assumption that noises among neighboring pixels are independent. However, this assumption is not applicable in LDCT since the noise in a CT image is known to be correlative. For LDCT denoising, Noise2Sim (Niu *et al* 2022) was proposed to denoise the correlative and structural noises by training with adjacent slices in the same 3D CT image. Noise2Inverse (Hendriksen *et al* 2020) splits the sinogram into several equal parts in terms of the projection angles and then reconstructs them into paired noise-independent images for network training. Although these methods have demonstrated promising results, they are essentially pure post-processing methods that operate only in the image domain and make no use of the sinogram. In fact, as more original measurements, sinograms contain lots of useful priors, abandoning the use of sinograms often means losing them. Hence, the utilization of

sinograms in self-supervised LDCT denoising methods has begun to attract attention. To explore priors in the sinogram domain, Zhou *et al* utilized the similarity of sinogram data from adjacent projection angles for self-supervised denoising (Zhou *et al* 2022). Furthermore, Wagner and Pfaff *et al* proposed a dual-domain self-supervised method (Wagner *et al* 2023). They split the sinogram by projection angles, connected the sinogram domain and image domain with a reconstruction operator, and trained the entire network with the image-domain loss. Niu and Wang *et al* applied Noise2Sim to both the sinogram and image domains, connected them with a reconstruction operator, and used a dual-domain loss to train the network, resulting in SSDDNet (Niu *et al* 2022). These dual-domain methods effectively utilize sinograms and improve denoising performance based on the image post-processing methods.

Although existing dual-domain self-supervised LDCT denoising methods do consider the sinogram domain and obtain positive results, they treat sinograms as general images and overlook important properties of low-dose sinograms. In fact, the noise levels of pixels in a low-dose sinogram are often inhomogeneous due to the nature of the CT projection process (Zhu *et al* 2012), which results in blurring artifacts in the reconstructed image when treated by an isotropic global denoiser like classic CNNs. If this degradation occurs, it would be challenging to deal with it in the image domain for self-supervised denoising. Another issue is that the denoising strength in the sinogram domain must be well-controlled to avoid introducing new secondary artifacts in the reconstructed image. This is because each single sinogram pixel would produce a global effect in the reconstruction process. So, denoising the sinogram should be done carefully such that only noise is removed. Unfortunately, existing dual-domain self-supervised methods do not address these issues but instead utilize classic CNNs for sinogram-domain denoising directly. Consequently, their final output images might suffer from blurring artifacts or streaks.

In this paper, we propose SDBDNet, an end-to-end dual-domain self-supervised deep learning method for LDCT denoising that optimizes the sinogram-domain denoising by explicitly taking the important properties of the sinogram into consideration. For self-supervised learning, we split the sinogram into two equal and non-overlapping parts based on the position of detector cells to acquire training data with good similarity and uncorrelated noise. Splitting the sinogram will significantly reduce the amount of data that can be used for reconstruction, which would introduce over-smoothing and artifacts in the reconstructed image. To tackle this problem, we use one-dimensional (1D) interpolation and learning-based correction to restore the sub-sinograms to their original full size. To tackle the inhomogeneous noise levels in the sinogram, our SDBDNet incorporates Dropblock layers to adaptively localize the effect of convolution for denoising, thus effectively reducing the blurring artifacts. In addition, to control the denoising strength of the sinogram domain and balance the effects of the sinogram-domain and image-domain denoising on the final result, a weighted averaging strategy to balance the denoised and noisy sinograms is proposed to further reduce possible artifacts due to over-smoothing.

Our work can be summarized as follows:

- We propose an end-to-end dual-domain self-supervised neural network model for LDCT denoising, which just assumes the availability of a representable dataset consisting of low-dose sinograms. For the first time, a weighting strategy is explicitly built into the proposed model to balance the behaviors of sinogram-domain denoiser and image-domain denoiser.
- Different from most other self-supervised LDCT denoising methods, we attach great importance to sinogram-domain denoising and make effective improvements. For the first time, we propose to incorporate the noise properties of low-dose sinograms into designing the network architecture in dual-domain self-supervised methods, which helps to avoid introducing secondary artifacts.
- Through numerical experiments, we demonstrate that our SDBDNet achieves superior performance in LDCT denoising compared to some traditional and self-supervised deep-learning denoising methods.

2. Related work

2.1. Image-domain methods

Unsupervised learning is a convenient and widely used approach for image denoising, with N2N (Lehtinen *et al* 2018) being a prominent example. The idea of N2N has been extended to the field of LDCT denoising, e.g. Noise2Sim (Niu *et al* 2022) and Noise2Inverse (Hendriksen *et al* 2020), and achieved promising results.

Noise2Noise proposes a denoising method that uses pairs of noisy images of the same scene without clean labels, under the assumption of zero-mean and uncorrelated noise. N2N has demonstrated that its performance is comparable to supervised methods when the number of training samples is large enough. N2N trains networks using noisy paired images to supervise each other:

$$\theta^* = \operatorname{argmin}_{\theta} \frac{1}{N} \sum_i \|\mathcal{F}(\mathbf{x}_i + \mathbf{n}_{i1}; \theta) - (\mathbf{x}_i + \mathbf{n}_{i2})\|_2^2, \quad (1)$$

where \mathbf{x}_i is the ideal clean signal of i th training image, \mathbf{n}_{i1} and \mathbf{n}_{i2} are two independent noise distributions. The function \mathcal{F} is the denoising neural network (NN) parameterized by θ , and N is the total number of the training samples.

Although N2N eliminates the need for labeled training images, acquiring two different noisy observations of the same scene, particularly in CT, can be also difficult. To address this issue, Noise2Sim (Niu *et al* 2022) replaces the needed paired noisy images with neighboring slices chosen from the same 3D CT image. The denoising network is then trained by minimizing the following loss:

$$\theta^* = \operatorname{argmin}_{\theta} \frac{1}{N} \sum_i \|\mathcal{F}(\mathbf{x}_i + \mathbf{n}_i; \theta) - (\mathbf{x}_i + \boldsymbol{\delta}_i + \hat{\mathbf{n}}_i)\|_2^2, \quad (2)$$

where \mathbf{n}_i and $\hat{\mathbf{n}}_i$ are two different noise distributions. $\boldsymbol{\delta}_i$ is the difference between the uncontaminated content in the two similar image slices. Noise2Sim does not require two noisy images of the same scene, but the need for neighboring slices makes it not a strict self-supervised method.

According to the characteristics of CT reconstruction, Allard proposed a strict self-supervised method Noise2Inverse (Hendriksen *et al* 2020) for LDCT denoising by generating multiple noisy observations from the same sinogram. Defining an operator $\mathcal{S}_v: R^{N_v \times N_d} \rightarrow (R^{(N_v/n) \times N_d}, R^{(N_v/n) \times N_d}, \dots, R^{(N_v/n) \times N_d})$ that maps a $N_v \times N_d$ object (e.g. a sinogram) into n non-overlapping sub-objects of the same size, Noise2Inverse splits a low-dose sinogram \mathbf{s}_i into n sub-sinograms by applying \mathcal{S}_v :

$$(\mathbf{s}_{i1}, \mathbf{s}_{i2}, \dots, \mathbf{s}_{in}) = \mathcal{S}_v(\mathbf{s}_i), \quad (3)$$

where N_v and N_d are the number of projection angles and detector cells, respectively. With a reconstruction operator $\mathcal{R}: R^{(N_v/n) \times N_d} \rightarrow R^{N_v \times N_d}$ (e.g. FBP or SART), each sub-sinogram \mathbf{s}_{ij} is reconstructed into a $N_v \times N_d$ sized noisy image $\hat{\mathbf{x}}_{ij}$:

$$\hat{\mathbf{x}}_{ij} = \mathcal{R}(\mathbf{s}_{ij}). \quad (4)$$

Since there are no explicit correlations between the sub-sinograms, the resulting images can be treated as multiple observations of the same scene with uncorrelated noise. Similar to N2N, Noise2Inverse trains networks with one image as the label and the average of the other images as the input:

$$\theta^* = \operatorname{argmin}_{\theta} \frac{1}{N} \sum_i \left\| \mathcal{F} \left(\frac{1}{n-1} \sum_{j=1}^{n-1} \hat{\mathbf{x}}_{ij}; \theta \right) - \hat{\mathbf{x}}_{in} \right\|_2^2. \quad (5)$$

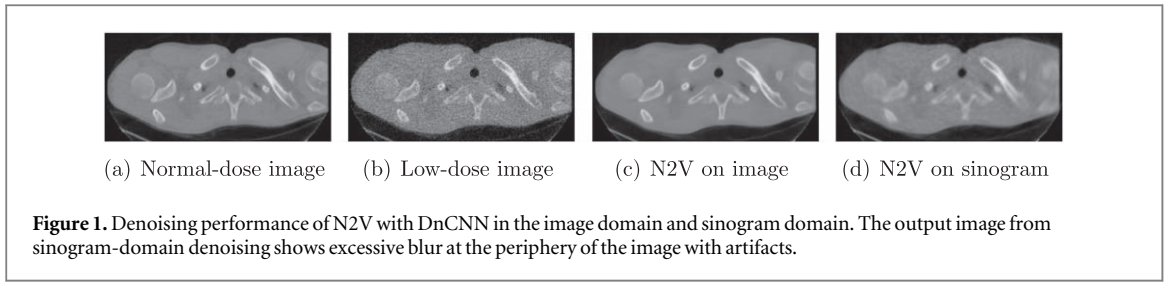
Noise2Inverse shows an effective way to obtain different noisy observations by sinogram-splitting. However, splitting the sinogram into several parts will result in blurring or even sparse-angle artifacts in the reconstructed images as only half or even less data are used for each separate reconstruction.

Although the aforementioned image-domain methods exhibit effective denoising capabilities, they disregard the valuable priors in the sinogram domain. Without the participation of the sinogram, the use of information existing in the measurement could be inadequate, and it might also lead to inconsistencies between the reconstructed image and its corresponding sinogram during the denoising process, resulting in excessive smoothness and blurring.

2.2. Dual-domain methods

In order to achieve more efficient LDCT denoising, self-supervised methods concerning the sinogram domain have been proposed. Zhou *et al* (2022) employed the sinogram data of adjacent projection angles to supervise each other for self-supervised denoising. Wagner *et al* (2023) employed a method to split each sinogram into two parts and set up networks to denoise the sub-sinograms. Through an embedded operator, the denoised sub-sinograms are reconstructed into images for self-supervised denoising, similar to what Noise2Inverse does. Niu *et al* extended Noise2Sim to the sinogram domain and proposed SSDDNet (Niu *et al* 2022). They used adjacent slices to supervise each other in both the sinogram and image domains. With the utilization of priors in the sinogram domain, these dual-domain methods achieve good denoising performance. However, the image blurring caused by sinogram splitting has not been addressed in these dual-domain methods, and denoising with adjacent slices is still not strictly self-supervised.

While sinogram-domain denoising has gained attention in the aforementioned dual-domain methods, they all employ denoising strategies designed for general images, overlooking the unique noise characteristics of low-dose sinograms. In fact, sinograms tend to suffer from non-stationary noise distribution (Deng *et al* 2019). Global isotropic denoisers like Gaussian filters and classical CNNs lead to blurring and artifacts easily when applied for sinogram-domain denoising (Zhu *et al* 2012). Figure 1(d) shows the reconstructed image with a



sinogram denoised by a typical self-supervised method Noise2Void (Krull *et al* 2019) embedded with a classical deep learning denoiser DnCNN (Zhang *et al* 2017), where the blurring artifacts are obvious when compared to the result by applying the same method in the image domain as shown in figure 1(c). It is well-known that each pixel in the sinogram has a global influence on the reconstruction process, and the quality of the reconstructed image is highly sensitive to small changes in the sinogram. Thus, effectively managing the denoising intensity in the sinogram domain becomes paramount. However, existing dual-domain self-supervised denoising methods do not take this into consideration. In contrast, our SDBDNet employs Dropblock layers to achieve adaptive denoising of the non-uniform noise levels in the low-dose sinogram, significantly reducing the blurring artifacts in the reconstructed image. Additionally, SDBDNet employs an explicit weighted averaging process to balance the influence of the sinogram-domain and image-domain denoising, effectively controlling the denoising strength in the sinogram domain, thereby further reducing the over-smoothing problem.

3. Method

SDBDNet consists of essentially three ingredients: the sinogram-domain network, the reconstruction layer, and the image-domain network. We perform self-supervised denoising in both the sinogram domain and the image domain, and incorporate the Ordered-subset Simultaneous Algebraic Reconstruction Technique (OS-SART) (Andersen and Kak 1984) as the reconstruction layer. This section will introduce these three ingredients in detail, and illustrate the overall pipeline and the algorithm flow-chart of SDBDNet.

3.1. The sinogram-domain network

In the sinogram domain, we achieve self-supervised denoising by mutual supervision of the low-dose data corresponding to adjacent detector cells, and since they come from different rays, their noise distributions are uncorrelated. Besides, the noise approximately follows a non-stationary Gaussian distribution (Zhu *et al* 2012). Therefore, we can use adjacent pixels along the detector cells for self-supervised denoising since the basic assumptions of the N2N (Lehtinen *et al* 2018) are met. By extracting the sinogram at equal intervals and splitting it into two parts, we can obtain a pair of sub-sinograms with similar content and uncorrelated noise, which can be used for training the self-supervised denoising network. Compared to the splitting strategy adopted in the literature that splits a sinogram according to the projection angles, we choose a more rational approach: we use the sinogram data from all projection angles while splitting the sinogram into two equal parts in terms of the positions of detector cells. It is known that the sinogram data has stronger neighboring correlations along the detector cells rather than along the projection angles (Wang *et al* 2006), thus our splitting strategy should lead to a better pair of noisy observations. Mathematically, defining an operator $\mathcal{S}_d: R^{N_v \times N_d} \rightarrow (R^{N_v \times (N_d/2)}, R^{N_v \times (N_d/2)})$ that splits a $N_v \times N_d$ sized sinogram into two equal parts according to the parity of the indices of the detector cells, we have:

$$(\mathbf{s}_{i,o}, \mathbf{s}_{i,e}) = \mathcal{S}_d(\mathbf{s}_i), \quad (6)$$

where \mathbf{s}_i is the noisy sinogram, $\mathbf{s}_{i,o}$ and $\mathbf{s}_{i,e}$ are the two obtained sub-sinograms, respectively.

To address the sinogram halving issue that arises from splitting, we utilize interpolation and a learning-based correction to inpaint the missing sinogram data, similar to some widely used approaches for inpainting sparse-angle CT sinograms. With a 1D linear interpolation operator $\mathcal{I}: R^{N_v \times (N_d/2)} \rightarrow R^{N_v \times N_d}$ that along the direction of the data corresponding to each projection angle (rows of the sinogram), as shown in figure 2, we patch the half-sized sub-sinograms to the full size while keeping the existing data unchanged:

$$\begin{aligned} \mathbf{s}'_{i,o} &= \mathcal{I}(\mathbf{s}_{i,o}), \\ \mathbf{s}'_{i,e} &= \mathcal{I}(\mathbf{s}_{i,e}). \end{aligned} \quad (7)$$

Then, we perform data correction and self-supervised denoising on $\mathbf{s}'_{i,o}$ and $\mathbf{s}'_{i,e}$ through the sinogram-domain network \mathcal{F}_S which is parameterized by θ_S . We feed $\mathbf{s}'_{i,o}$ and $\mathbf{s}'_{i,e}$ into the network and make each other as

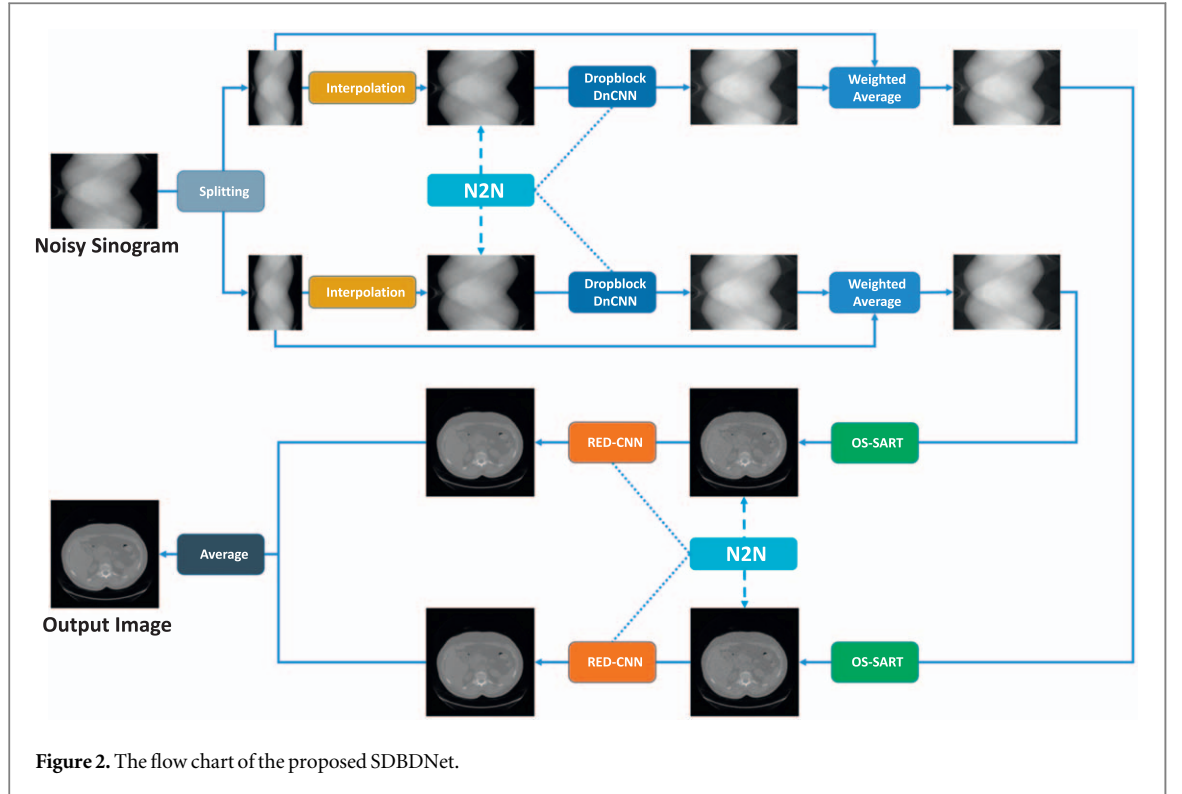


Figure 2. The flow chart of the proposed SDBDNet.

their labels to train the network:

$$\theta_S^* = \operatorname{argmin}_{\theta_S} \frac{1}{N} \sum_i (\|\mathcal{F}_S(\mathbf{s}'_{i,o}; \theta_S) - \mathbf{s}'_{i,e}\|_2^2 + \|\mathcal{F}_S(\mathbf{s}'_{i,e}; \theta_S) - \mathbf{s}'_{i,o}\|_2^2). \quad (8)$$

By introducing the mean squared error (MSE) loss function represented by \mathcal{L}_{mse} , we can import the loss function \mathcal{L}_{Sino} of the sinogram-domain denoising network from (8):

$$\mathcal{L}_{Sino} = \mathcal{L}_{mse}(\mathcal{F}_S(\mathbf{s}'_{i,o}; \theta_S), \mathbf{s}'_{i,e}) + \mathcal{L}_{mse}(\mathcal{F}_S(\mathbf{s}'_{i,e}; \theta_S), \mathbf{s}'_{i,o}). \quad (9)$$

Thus, the denoised sinograms $\mathbf{s}_{i,1}$ and $\mathbf{s}_{i,2}$ is given by:

$$\begin{aligned} \mathbf{s}_{i,1} &= \mathcal{F}_S(\mathbf{s}'_{i,o}; \theta_S), \\ \mathbf{s}_{i,2} &= \mathcal{F}_S(\mathbf{s}'_{i,e}; \theta_S). \end{aligned} \quad (10)$$

With the above training strategy, the network can effectively perform data correction and unsupervised denoising simultaneously. In fact, the learning process of the network on the real sinogram data $\mathbf{s}_{i,k}^{real}$ corresponding to the k_{th} detector cell can be expressed as:

$$\begin{aligned} \theta^* &= \operatorname{argmin}_{\theta} \|\mathcal{F}_S(\mathbf{s}_{i,k}^{real}; \theta) - \mathbf{s}_{i,k}^{intp}\|_2^2 \\ &= \operatorname{argmin}_{\theta} \left\| \mathcal{F}_S(\mathbf{s}_{i,k}^{real}; \theta) - \frac{1}{2}(\mathbf{s}_{i,k-1}^{real} + \mathbf{s}_{i,k+1}^{real}) \right\|_2^2. \end{aligned} \quad (11)$$

While on the interpolated data $\mathbf{s}_{i,k}^{intp}$ corresponding to the k_{th} detector cell can be expressed as:

$$\theta^* = \operatorname{argmin}_{\theta} \|\mathcal{F}_S(\mathbf{s}_{i,k}^{intp}; \theta) - \mathbf{s}_{i,k}^{real}\|_2^2. \quad (12)$$

Equation (11) demonstrates that when acting on real sinogram data, the network learns to perform self-supervised denoising, while equation (12) shows that when acting on interpolated data, the network learns to correct the interpolated data. Consequently, the trained sinogram-domain network can effectively perform both sinogram correction and denoising simultaneously.

As mentioned earlier, CNN denoisers with conventional convolution kernels have difficulties in dealing with inhomogeneous noise levels. One solution to this problem is to modify the global action of convolution kernels

and adjust the denoising strength in an adaptive manner. In our study, we replaced the traditional convolutional layer in the sinogram-domain network with the Dropblock layer (Ghiasi *et al* 2018). The Dropblock layer has been successful in image segmentation and recognition as a regularization method for convolutional networks. It performs dropout on the feature produced by each convolution kernel. When used for sinogram-domain denoising, the Dropblock layer discards part of the effect of each convolution kernel with a drop rate r instead of acting on the entire sinogram. As a result, the denoising effect of each kernel is limited to only a portion of the sinogram. This allows each convolution kernel to denoise only sinogram regions where the noise levels are more similar, effectively alleviating the problem of excessive denoising caused by global denoisers.

3.2. The OS-SART layer

To construct a dual-domain method, we need to reconstruct sinograms into images with a reconstruction operator $\mathcal{R}: R^{N_v \times N_d} \rightarrow R^{N \times N}$. We select the ordered-subset simultaneous algebraic reconstruction technique (OS-SART) (Andersen and Kak 1984) as our reconstruction operator, since it is commonly used in the CT community. We incorporate OS-SART into the dual-domain network as a layer to connect the sinogram and image domains. After randomly permutating the reconstruction angles, we found that the OS-SART algorithm with 1 iteration could reconstruct an image of high quality, and more iterations would not bring significant improvements. Hence, we set the iteration number of the OS-SART layer to 1. In fact, more iterations will significantly increase the calculations of backpropagation when training the network, thereby increasing the training time and storage requirements. In this case, an iteration number of 1 is a rational choice.

3.3. The image-domain network

In the image domain, the two images reconstructed from the denoised sinograms are used as each other's labels to train the denoising network. However, instead of directly utilizing the sinograms output from the sinogram-domain network, we opt for a weighted average of the denoised sinograms and their corresponding original noisy ones. The primary objective of this weighted average is to further control the sinogram-domain denoising strength and restore some details that may have been blurred during sinogram-domain denoising. Although this reintroduces some noise, the noise level is significantly reduced compared to that without sinogram-domain denoising. Moreover, the reintroduced noise undergoes processing in the image-domain network, which should be again effectively removed.

The weights are calculated based on the column sum of the original noisy sinogram. It is known that larger sinogram values correspond to higher noise levels (Deng *et al* 2019). As a simple strategy, we utilize the column sums of the sinogram as an indicator of the magnitude of noise levels presented in each column. To obtain the weights, we first sum the pixel values on each column of the original noisy sinogram \mathbf{s}_i with an operator $\mathcal{M}: R^{N_v \times N_d} \rightarrow R^{N_d}$, normalize it with its maximum value, and then use a suitable scalar $\beta \in [0, 1]$ to perform scaling of the weight. In detail, the resulting weight vector is:

$$\boldsymbol{\alpha} = \frac{\beta^* \mathcal{M}(\mathbf{s}_i)}{\max(\mathcal{M}(\mathbf{s}_i))}, \quad (13)$$

where \max is the operator to compute the maximum value of a vector. In order to correctly multiply $\boldsymbol{\alpha}$ with the corresponding projection values in the sinogram, we stretch out $\boldsymbol{\alpha}$ to size $N_v \times N_d$, with a row copy operator $\mathcal{C}: R^{N_d} \rightarrow R^{N_v \times N_d}$, and define $N_v \times N_d$ sized odd-even indicator matrices I_o and I_e , where I_o takes the value 1 in the odd columns and 0 in the even columns while $I_e = 1 - I_o$ is the opposite. In the weighted average, $\boldsymbol{\alpha}$ corresponds to the weight of the denoised sinograms $\mathbf{s}_{i,1}$ and $\mathbf{s}_{i,2}$, and $(1 - \boldsymbol{\alpha})$ corresponds to the original noisy one \mathbf{s}_i . Thus, the weighted sinograms $\bar{\mathbf{s}}_{i,1}$ and $\bar{\mathbf{s}}_{i,2}$ are given by:

$$\begin{aligned} \bar{\mathbf{s}}_{i,1} &= (I_o^* \mathcal{C}(\boldsymbol{\alpha}) + I_e) * \mathbf{s}_{i,1} + I_o^* \mathcal{C}(1 - \boldsymbol{\alpha}) * \mathbf{s}_i, \\ \bar{\mathbf{s}}_{i,2} &= (I_e^* \mathcal{C}(\boldsymbol{\alpha}) + I_o) * \mathbf{s}_{i,2} + I_e^* \mathcal{C}(1 - \boldsymbol{\alpha}) * \mathbf{s}_i. \end{aligned} \quad (14)$$

Once the two averaged sinograms are obtained, we use the OS-SART layer (represented by \mathcal{R}) to reconstruct them into images $\mathbf{x}_{i,o}$ and $\mathbf{x}_{i,e}$:

$$\begin{aligned} \mathbf{x}_{i,o} &= \mathcal{R}(0, \bar{\mathbf{s}}_{i,1}, \omega), \\ \mathbf{x}_{i,e} &= \mathcal{R}(0, \bar{\mathbf{s}}_{i,2}, \omega), \end{aligned} \quad (15)$$

where 0 is the $N \times N$ sized initial image set as zero-values and ω represents the relaxation parameter of OS-SART. The resulting images are then fed into the image-domain network as labels for each other to train the denoising network \mathcal{F}_I parameterized by θ_I :

$$\theta_I^* = \operatorname{argmin}_{\theta_I} \frac{1}{N} \sum_i (\|\mathcal{F}_I(\mathbf{x}_{i,o}; \theta_I) - \mathbf{x}_{i,e}\|_2^2 + \|\mathcal{F}_I(\mathbf{x}_{i,e}; \theta_I) - \mathbf{x}_{i,o}\|_2^2). \quad (16)$$

Algorithm 1. SDBDNet

Input: The noisy sinogram \mathbf{s}_i , initial image $\mathbf{x}_i^{(0)} = \mathbf{0}$, the odd–even indicator matrices I_o and I_e , the training iteration number n and the relaxation parameter ω for OS-SART.

Operator: $\mathcal{S}_d, \mathcal{I}, \mathcal{C}, \mathcal{R}$.

Networks: \mathcal{F}_S parameterized by θ_S , \mathcal{F}_I parameterized by θ_I .

Output: The denoised image $\mathbf{x}_{i,out}$.

Training:

for $k = 0; k < n; k++$ do

for every \mathbf{s}_i in the training set

(1) Splitting the sinogram \mathbf{s}_i into two parts, by:

$$(\mathbf{s}_{i,o}, \mathbf{s}_{i,e}) = \mathcal{S}_d(\mathbf{s}_i).$$

(2) Interpolating $\mathbf{s}_{i,o}, \mathbf{s}_{i,e}$ to the full size, by:

$$\mathbf{s}'_{i,o} = \mathcal{I}(\mathbf{s}_{i,o}),$$

$$\mathbf{s}'_{i,e} = \mathcal{I}(\mathbf{s}_{i,e}).$$

(3) Correcting and denoising $\mathbf{s}'_{i,o}$ and $\mathbf{s}'_{i,e}$ with the sinogram-domain network:

$$\mathbf{s}_{i,1} = \mathcal{F}_S(\mathbf{s}'_{i,o}; \theta_S),$$

$$\mathbf{s}_{i,2} = \mathcal{F}_S(\mathbf{s}'_{i,e}; \theta_S).$$

(4) Solving the weighted average of the sinograms:

$$\bar{\mathbf{s}}_{i,1} = (I_o * \mathcal{C}(\alpha) + I_e) * \mathbf{s}_{i,1} + I_o * \mathcal{C}(1 - \alpha) * \mathbf{s}_i,$$

$$\bar{\mathbf{s}}_{i,2} = (I_e * \mathcal{C}(\alpha) + I_o) * \mathbf{s}_{i,2} + I_e * \mathcal{C}(1 - \alpha) * \mathbf{s}_i.$$

(5) Reconstructing images using the OS-SART layer:

$$\mathbf{x}_{i,o} = \mathcal{R}(0, \bar{\mathbf{s}}_{i,1}, \omega),$$

$$\mathbf{x}_{i,e} = \mathcal{R}(0, \bar{\mathbf{s}}_{i,2}, \omega).$$

(6) Denoising the images with the image-domain network:

$$\mathbf{x}_{i,1} = \mathcal{F}_I(\mathbf{x}_{i,o}; \theta_I),$$

$$\mathbf{x}_{i,2} = \mathcal{F}_I(\mathbf{x}_{i,e}; \theta_I).$$

(7) Training the dual-domain denoising network with the dual-domain loss:

$$\mathcal{L}_{Dual} = \mathcal{L}_{Sino} + \lambda * \mathcal{L}_{Img}.$$

end for

end for

Testing:

(1) Repeating the above processes (1)–(6) in the training stage with the well-trained networks \mathcal{F}_S parameterized by θ_S^* and \mathcal{F}_I parameterized by θ_I^* .

(2) Taking the average of the two denoised images to obtain the final denoised image result.

$$\mathbf{x}_{i,out} = \frac{1}{2}(\mathbf{x}_{i,1} + \mathbf{x}_{i,2}).$$

The loss of the image-domain networks \mathcal{L}_{Img} is given by:

$$\mathcal{L}_{Img} = \mathcal{L}_{mse}(\mathcal{F}_I(\mathbf{x}_{i,o}; \theta_I), \mathbf{x}_{i,e}) + \mathcal{L}_{mse}(\mathcal{F}_I(\mathbf{x}_{i,e}; \theta_I), \mathbf{x}_{i,o}). \quad (17)$$

Thus, the denoised images $\mathbf{x}_{i,1}$ and $\mathbf{x}_{i,2}$ are given as:

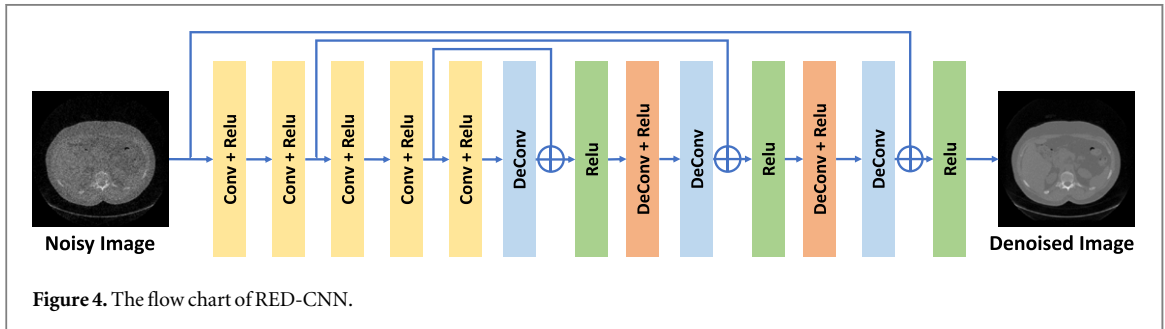
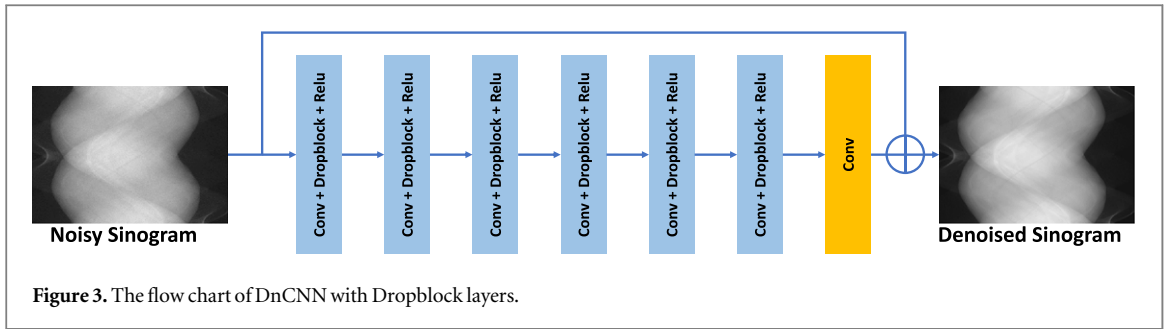
$$\mathbf{x}_{i,1} = \mathcal{F}_I(\mathbf{x}_{i,o}; \theta_I),$$

$$\mathbf{x}_{i,2} = \mathcal{F}_I(\mathbf{x}_{i,e}; \theta_I). \quad (18)$$

To combine the losses of the sinogram-domain and image-domain networks, we introduce a hyper-parameter λ as the weighting parameter. The total loss \mathcal{L}_{Dual} of the dual-domain network is given by:

$$\mathcal{L}_{Dual} = \mathcal{L}_{Sino} + \lambda * \mathcal{L}_{Img}. \quad (19)$$

During the testing phase, we compute the average of the output images from \mathcal{F}_I to obtain the final denoising result $\mathbf{x}_{i,out}$ as expressed by:



$$\mathbf{x}_{i,out} = \frac{1}{2}(\mathbf{x}_{i,1} + \mathbf{x}_{i,2}). \quad (20)$$

3.4. Pipeline

Our proposed SDBDNet is an end-to-end solution that produces the denoised image directly from the noisy sinogram, as illustrated in figure 2. In the sinogram domain, we employ a 7-layer DnCNN (Zhang *et al* 2017) (shown in figure 3), with Dropblock (Ghiasi *et al* 2018) layers replacing the standard convolutional layers. In the image domain, we utilize the network of RED-CNN (Chen *et al* 2017), which has shown excellent performance on LDCT image denoising, as depicted in figure 4. The entire denoising process is outlined in algorithm 1.

4. Experiments

In this section, we will evaluate the performance of our SDBDNet through experiments and ablation studies.

Our experiments were performed on two datasets: (1) the ‘2016 NIH-AAPM-Mayo Clinic Low Dose CT Grand Challenge dataset’ (AAPM dataset) (McCullough 2016) and (2) the ‘LoDoPaB dataset’ (Leuschner *et al* 2021). For both datasets, we utilized the normal-dose CT images as the ground truth and generated the clean sinograms using a simulated CT algorithm with a fan beam source and 360 projection angles uniformly distributed in the angular range $[0, 2\pi]$, along with a linear detector consisting of 512 cells. To simulate radiation with different doses, Poisson noise was added to the raw data, given by

$$p_n = -\ln\left(\frac{I_d}{I_0}\right), \quad I_d \sim \text{Poisson}\{I_0 \times e^{-p}\} \quad (21)$$

where p and p_n represent the clean and noisy sinograms, respectively. The symbol I_0 indicates the number of incident photons, and I_d represents the number of photons collected by the detector. In our experiments, we set $I_0 = 1 \times 10^6$ as the normal dose, while $I_0 = 1 \times 10^4$ (1% of the normal dose) and $I_0 = 0.5 \times 10^4$ (0.5% of the normal dose) as two low-dose conditions. For the proposed SDBDNet, we split the sinograms into two non-overlapping, uniformly distributed groups based on the parity of the position of the detector cells within a range of $[1, 512]$, following the depiction in figure 2.

We reconstructed the normal-dose sinograms into images directly using the OS-SART algorithm and made them the reference images for evaluating the denoising performance of the proposed SDBDNet and other comparison methods. To assess the quality of the output images, we chose the peak signal-to-noise ratio (PSNR) (Hore and Ziou 2010) and structural similarity index measure (SSIM) (Hore and Ziou 2010) as our evaluation

metrics. For comparison, we computed the average PSNR and SSIM of all results on the test set for each method. As a benchmark, we also directly reconstructed low-dose images from the test set and calculated their scores.

In terms of the choice of comparison methods, we chose two traditional well-known denoising methods: NLM filter (Buades *et al* 2005) and BM3D filter (Dabov *et al* 2007). NLM and BM3D are two gold standards for performance evaluation in image denoising, although the noise of LDCT images does not fully comply with their Gaussian-noise assumption, their results still have reference values. Here we use their results as a benchmark and reference. In learning-based methods, we chose two state-of-the-art (SOTA) self-supervised learning-based denoising methods: Noise2Void (Krull *et al* 2019) (N2V) and Noise2Inverse (Hendriksen *et al* 2020), a recently proposed dual-domain self-supervised denoising method (Wagner *et al* 2023), referred as ETSRP (end-to-end trainable and self-supervised CT reconstruction pipeline, described by the original author), and a well-known supervised LDCT denoising method RED-CNN (Chen *et al* 2017). For the Noise2Inverse method, there is a choice for the number of sub-sinograms. We tested two typical cases: N2Inverse-v2 and N2Inverse-v4, respectively, for the sinogram to be split into 2 and 4 sub-sinograms.

Python codes for all the networks referred to above are publicly available. Also, the NLM filter and BM3D filter were conducted by publicly available Python code in the SciPy library (<https://scipy.org/>) and BM3D library (<http://webpages.tuni.fi/foi/GCF-BM3D/>), respectively. During network training, we utilized the Mean Squared Error (MSE) loss function and the Adam (Kingma and Ba 2014) algorithm with an initial learning rate of 0.0001 for optimization. The mini-batch size was set to 8, and the network was trained for 1000 epochs, *stopping when loss reduction was no longer significant for several epochs*. All experiments were conducted on a server running Ubuntu 18.04, with Python 3.9.0, PyTorch 1.2.0, and a Nvidia RTX 2080Ti GPU card. The MSE loss function and Adam algorithm were provided by the PyTorch library, whereas the OS-SART algorithm was coded using CUDA kernels wrapped by the Cupy library (<https://github.com/cupy/cupy>).

4.1. Experiments on the AAPM dataset

The AAPM dataset (McCollough 2016) is a clinical dataset containing 1308 thoracic volumes from 1010 individuals, comprising 244 527 image slices with dimensions of 512×512 . From this dataset, we selected seven patients' 2209 slices as the training set, one patient's 340 slices for validation, and another patient's 344 slices for testing. We performed experiments on this dataset with two low-dose conditions: $I_0 = 1 \times 10^4$ (1%) and $I_0 = 0.5 \times 10^4$ (0.5%). After several trial and error tests, we set the parameters of the proposed SDBDNet as follows: the scaling factor of the weighted average $\beta = 0.5$, the drop rate of the Dropblock layers $r = 0.3$ and the weighting hyper-parameter of the dual-domain loss $\lambda = 0.5$, for both the conditions.

To verify the denoising effect in the sinogram domain, we directly output the results of SDBDNet in the sinogram domain (SD) and reconstructed them into images. We compared the results with that of the N2V method acting directly on the sinogram domain. Figure 5 presents the output sinograms and the reconstructed images of the experiments with the low-dose condition of $I_0 = 1 \times 10^4$ and $I_0 = 0.5 \times 10^4$, respectively. As shown in figures 5(c), (g), (k) and (o), N2V trends to over-denoise the sinogram and leads to loss of some lines in the sinogram. The structures in the reconstructed images exhibit severe blurring and adhesion, especially in the periphery. In contrast, the sinogram-domain network of SDBDNet, which uses the Dropblock layers instead of the traditional convolutional layers, can make effective denoising to a certain extent while avoiding blurring artifacts, as shown in figures 5(d), (h), (l) and (p). Please note that as a dual-domain method, SDBDNet pursues moderate denoising rather than complete denoising in the sinogram domain, such that structures are preserved as well as possible. Table 1 lists the quantitative measures, including average PSNR and SSIM, of each method with the low-dose conditions of $I_0 = 1 \times 10^4$ and $I_0 = 0.5 \times 10^4$, respectively. All average PSNR and SSIM values were calculated using the corresponding normal-dose CT images. The quantitative results demonstrate that the proposed SDBDNet offers clear advantages. In experiments with the low-dose condition of $I_0 = 1 \times 10^4$, SDBDNet achieves an average PSNR advantage of 0.89 db and an average SSIM advantage of 0.010 compared to other non-supervised methods. In experiments with the low-dose condition of $I_0 = 0.5 \times 10^4$, SDBDNet shows an average PSNR advantage of 1.04 db and an SSIM advantage of 0.013 compared to other non-supervised methods. Visual comparisons are presented in figures 6 and 7. Post-processing denoising methods that operate solely in the image domain are prone to losing structure details or generating pseudo-structures due to the lack of considering the sinogram. In contrast, as a dual-domain denoising method, SDBDNet presents clearer and more accurate image structure details. On the other hand, dual-domain methods with conventional CNN denoiser on the sinogram domain would introduce blurring artifacts because of the noise characteristics, which are not shown in the results of our SDBDNet due to our reasonable denoising control. In general, the image results demonstrate that SDBDNet can achieve outstanding denoising effects. Specifically, in figure 6(i), the elliptical structure pointed by the left arrow shows the shape closest to the reference, while other methods show excessive blur or deformation. Even the supervised method RED-CNN causes additional protrusion in the upper left corner of this structure. The crescent shadow structure pointed by the right arrow also shows the best

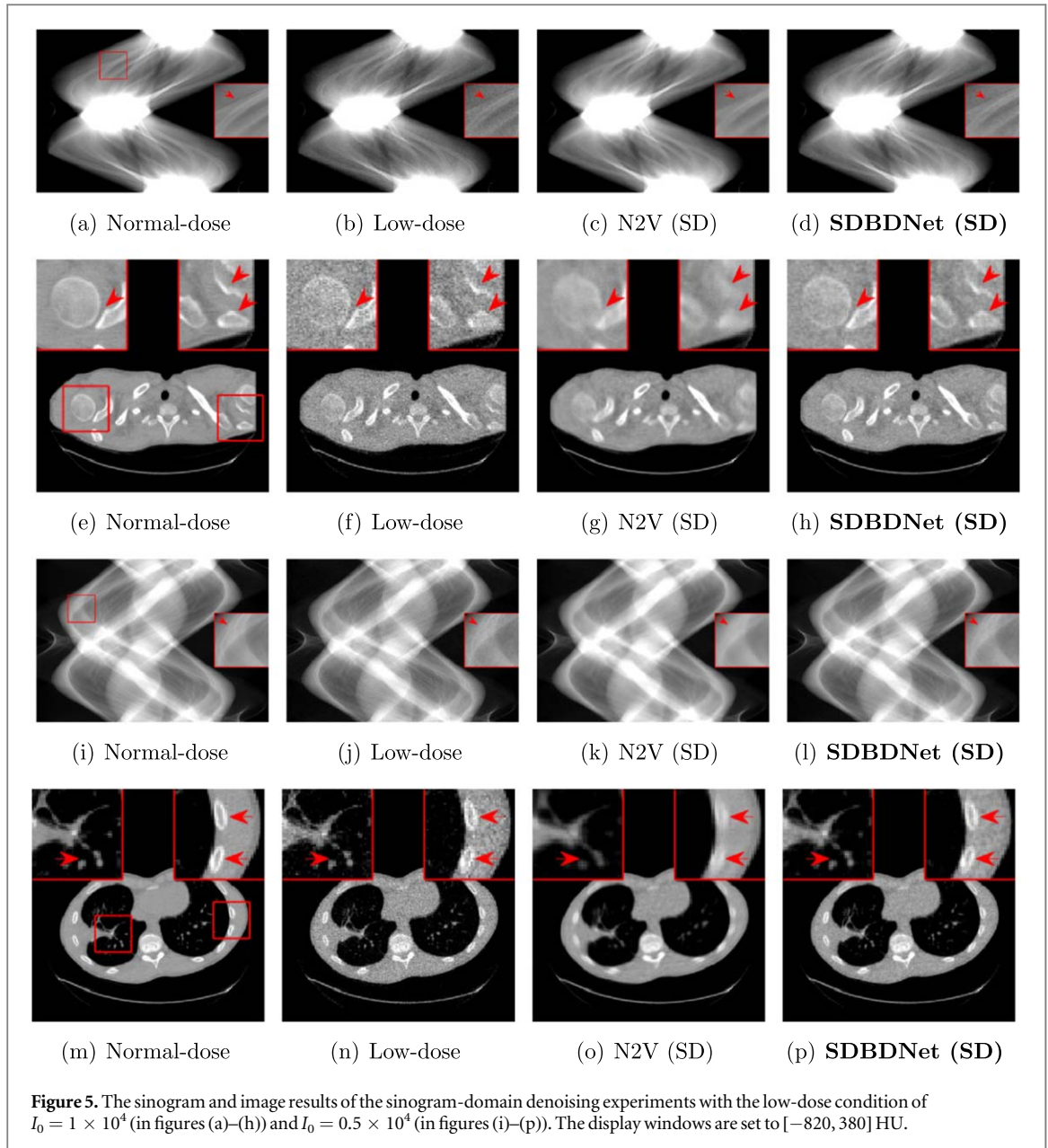
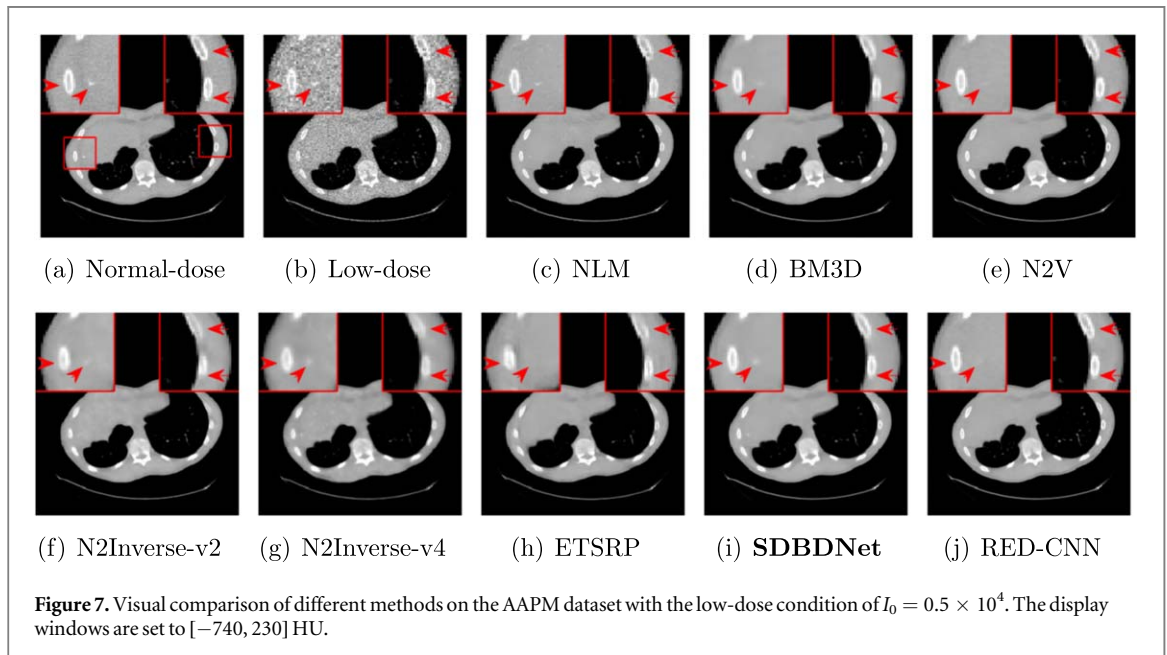
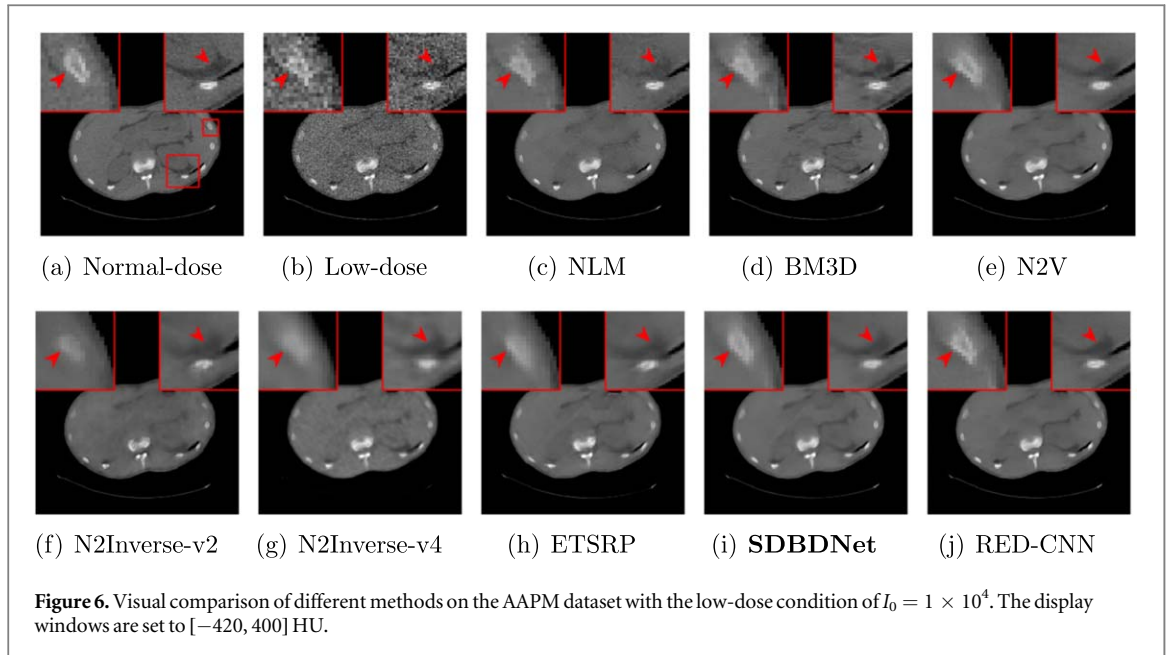


Table 1. comparison (psnr(db)/ssim) of different methods on the AAPM dataset with the low-dose condition of $I_0 = 1 \times 10^4$ (1%) and $I_0 = 0.5 \times 10^4$ (0.5%).

Method	$I_0 = 1 \times 10^4$		$I_0 = 0.5 \times 10^4$	
	PSNR	SSIM	PSNR	SSIM
Low-dose	31.75	0.793	29.18	0.724
NLM (Traditional)	35.95	0.899	35.13	0.880
BM3D (Traditional)	36.45	0.909	35.76	0.898
Noise2Void (Self)	36.81	0.917	35.83	0.901
Noise2Inverse-v2 (Self)	34.41	0.897	33.85	0.880
Noise2Inverse-v4 (Self)	33.44	0.883	32.91	0.864
ETSRP (Self+Dual)	31.25	0.876	30.92	0.864
Proposed SDBDNet	37.70	0.927	36.87	0.914
RED-CNN (Supervised)	38.04	0.932	37.20	0.921

shape restoration without forming adhesion with the cavity area below, while other comparison methods exhibit deformation and confusion. In figure 7(i), as shown in the zoomed-in area, SDBDNet restores the three elliptical structures with high quality while keeping the dot structure pointed by the second arrow on the left from



disappearing. In contrast, unsupervised learning methods N2V, ETSRP and even the supervised method RED-CNN failed to protect the dot structure. The results of both versions of N2Inverse exhibit high blurring. Although NLM shows higher sharpness at the dot structure, its effect on the three elliptical structures are bad.

4.2. Experiments on the LoDoPaB dataset

The LoDoPaB dataset (Leuschner *et al* 2021) is a simulated low-dose CT dataset that consists of over 40 000 out-of-order normal-dose CT image slices with a size of 362×362 pixels. In our experiments, we selected 2560 slices from this dataset, of which 2048 slices for training, 256 slices for validation, and 256 slices for testing. To ensure that the whole image was in the imaging field, we padded the original images to 512×512 pixels using zero-padding around them. We performed experiments with two low-dose conditions $I_0 = 1 \times 10^4$ (1%) and $I_0 = 0.5 \times 10^4$ (0.5%) on this dataset. For the parameters of the proposed SDBDNet, we set $\beta = 0.5$, $r = 0.2$, $\lambda = 1.0$ for experiments with the low-dose condition of $I_0 = 1 \times 10^4$ and $\beta = 0.5$, $r = 0.3$, $\lambda = 0.5$ for those with the low-dose condition of $I_0 = 0.5 \times 10^4$.

Table 2 presents the quantitative results (average PSNR and SSIM) of each method. The proposed SDBDNet demonstrates significant advantages in terms of these metrics. In experiments with the low-dose condition of $I_0 = 1 \times 10^4$, SDBDNet achieved an advantage of 1.11 db in average PSNR and 0.017 in average SSIM compared

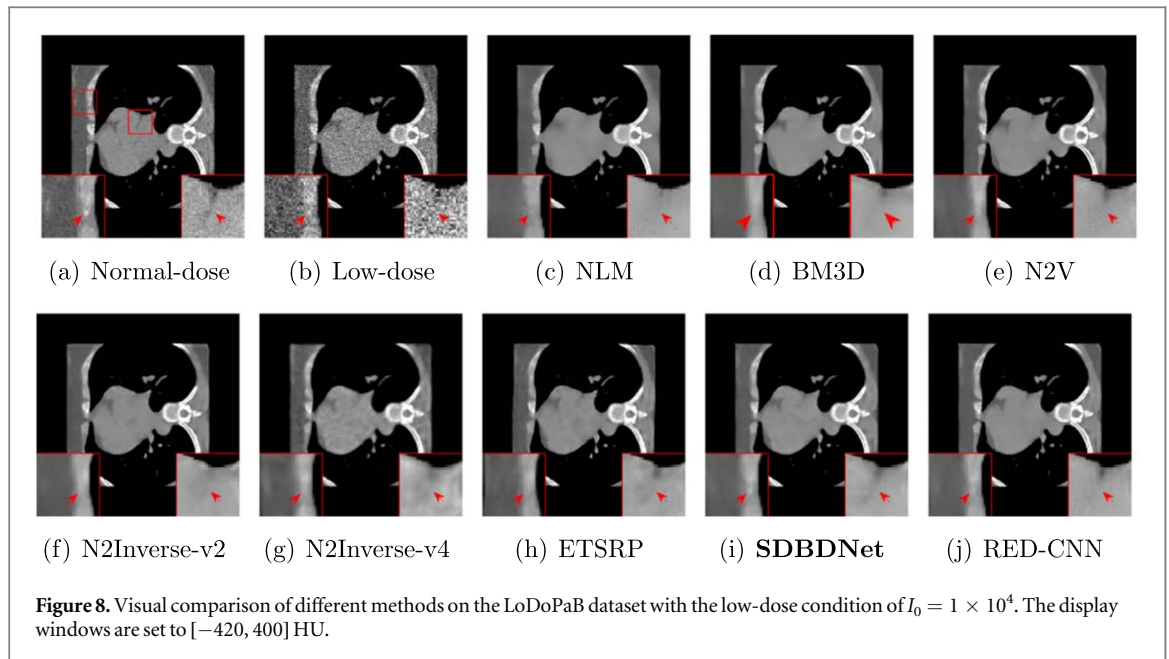


Table 2. Comparison (psnr(db)/ssim) of different methods on the LodoPaB dataset with the low-dose condition of $I_0 = 1 \times 10^4$ (1%) and $I_0 = 0.5 \times 10^4$ (0.5%).

Method	$I_0 = 1 \times 10^4$		$I_0 = 0.5 \times 10^4$	
	PSNR	SSIM	PSNR	SSIM
Low-dose	31.31	0.785	28.50	0.693
NLM (Traditional)	36.27	0.881	34.56	0.834
BM3D (Traditional)	37.47	0.914	35.92	0.880
Noise2Void (Self)	37.71	0.902	36.05	0.871
Noise2Inverse-v2 (Self)	36.69	0.891	34.60	0.801
Noise2Inverse-v4 (Self)	34.07	0.808	32.70	0.795
ETSRP (Self+Dual)	31.96	0.851	32.00	0.819
Proposed SDBDNet	38.82	0.931	37.45	0.898
RED-CNN (Supervised)	39.06	0.949	37.88	0.938

to other non-supervised methods. In experiments with the low-dose condition of $I_0 = 0.5 \times 10^4$, SDBDNet demonstrated an advantage of 1.40 db in average PSNR and 0.018 in average SSIM. The visual comparisons are presented in figures 8 and 9. As shown in the left zoomed-in area of figure 8(i), SDBDNet successfully retains the white dot structure pointed by the arrow while performing effective denoising, which has not been presented in other learning-based methods. In the right zoomed-in area, SDBDNet performs a good restoration of the shape and length of the curved shadow structure with low contrast in the image, showing the closest approximation to the reference image. As a contrast, even the supervised learning method RED-CNN failed to restore these two image details. In figure 9, SDBDNet also shows effective restoration to image structures. As shown in the above zoomed-in area of figure 9(i), while ensuring that the area pointed by the arrow is not excessively blurred, SDBDNet does not recognize the white-dot noise in the lower right corner as a structure and incorrectly retains it like N2V and RED-CNN. In the lower zoomed-in area, SDBDNet clearly restores the bladed structure indicated by the lower arrow without blurring the notch indicated by the upper arrow, outperforming all the other non-supervised comparison methods. Table 3 presents the comparison of each learning-based method in terms of the parameter size, training time and inference time in the experiments on the LodoPaB dataset with the low-dose condition of $I_0 = 1 \times 10^4$. As dual-domain methods, our proposed SDBDNet and ETSRP require more time for training and inference because of the dual-domain denoising and the embedded reconstructions. Since SDBDNet contains two reconstructions and an averaging, its inference time is more than double that of ETSRP. However, it is still at a fast level and will not show obvious disadvantages in applications. In terms of the amount of network parameters, SDBDNet does not rely on a large number of parameters, even less than the single-domain method Noise2Inverse.

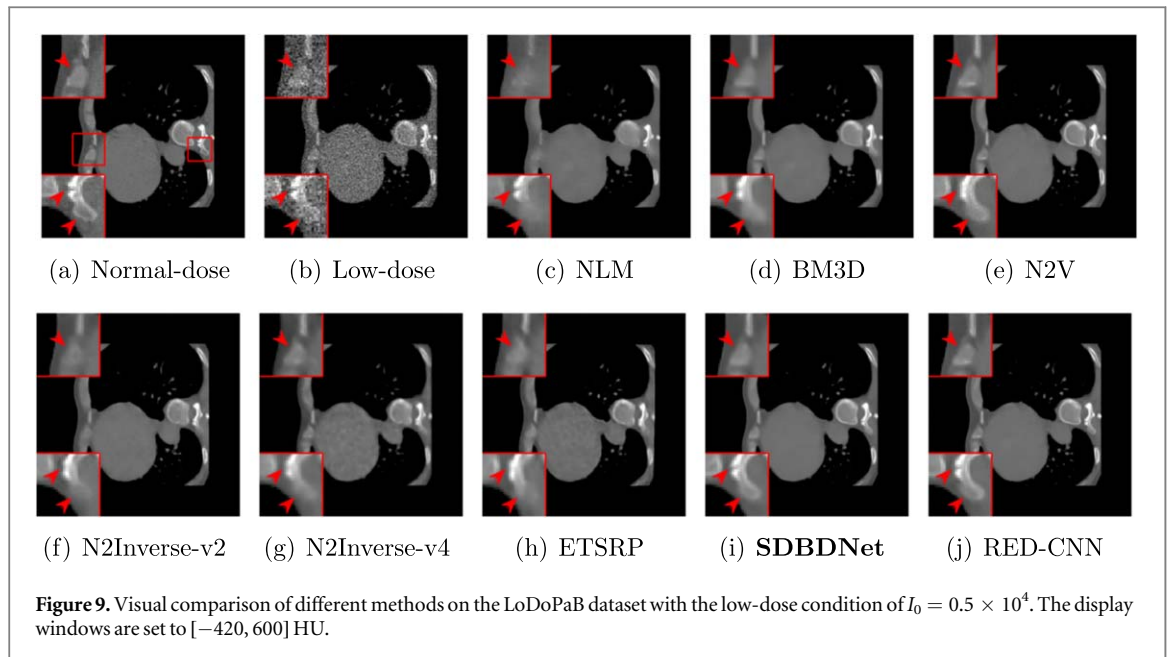


Table 3. The parameter size, training time and inference time of different learning-based methods in the experiments on the LodoPaB dataset with the low-dose condition of $I_0 = 1 \times 10^4$ (1%).

Method	Parameters	Training	Inference
Noise2Void (Self)	1.77 M	16 h	0.03 s
Noise2Inverse-v2 (Self)	2.58 M	13 h	0.02 s
Noise2Inverse-v4 (Self)	2.58 M	13 h	0.05 s
ETSRP (Self+Dual)	2.48 M	79 h	0.22 s
Proposed SDBDNet	2.48 M	56 h	0.51 s
RED-CNN (Supervised)	1.77 M	15 h	0.04 s

4.3. Analysis of experimental results

Experiments on the two datasets demonstrate that the proposed SDBDNet outperforms popular denoising methods including traditional non-learning methods and self-supervised learning-based methods, in LDCT denoising. The quantitative and visual results both confirm the effectiveness and improved performance of the proposed method.

Based on both quantitative and visual results, it is evident that the performance of Noise2Inverse and ETSRP is unsatisfactory in our experiments. We observed lower evaluation indices and poor image sharpness in these methods. Furthermore, splitting the sinogram into four parts resulted in worse denoising performance than splitting it into two parts for Noise2Inverse. This is because both Noise2Inverse and ETSRP use significantly less sinogram data to reconstruct images after splitting. In our experiments, the total number of imaging angles was 360, so that the sub-sinograms after the half split came from 180 angles, while the sub-sinograms after the four-part split came from 90 angles. In this case, the reconstructed images would be severely noisy and blurred, which negatively affected subsequent denoising results. These results validate the effectiveness of our interpolation and correction scheme for constructing sub-sinograms. On the other hand, ETSRP applies a classical CNN directly to denoise the sinogram without proper control, thereby worsening its performance due to the blurring from inappropriate sinogram denoising. This underscores the importance of our proposed SDBDNet's sinogram-domain denoising improvement.

Moreover, based on the visual results, dual-domain denoising is advantageous in restoring details compared to pure image-domain denoising. As demonstrated in figures 7–9, the proposed SDBDNet outperforms N2V and even the supervised method RED-CNN in the processing of tiny structures in images. This highlights the advantage of preserving details through dual-domain denoising.

Table 4. comparison (psnr(db)/ssim) of ablation experiments on the LodoPaB dataset with the low-dose conditions of $I_0 = 1 \times 10^4$ (1%) and $I_0 = 0.5 \times 10^4$ (0.5%).

Condition	$I_0 = 1 \times 10^4$		$I_0 = 0.5 \times 10^4$	
	PSNR	SSIM	PSNR	SSIM
Low-dose	31.31	0.785	28.50	0.693
$r = 0$	38.48	0.922	36.98	0.891
$\alpha \equiv 1$	37.53	0.922	36.58	0.888
$\alpha \equiv 0$	37.21	0.913	37.08	0.884
None BP	38.68	0.927	37.28	0.889
Proposed SDBDNet	38.82	0.931	37.45	0.898

4.4. Ablation experiments

We performed ablation experiments on the LoDoPaB dataset, with the low-dose conditions of $I_0 = 1 \times 10^4$ (1%) and $I_0 = 0.5 \times 10^4$ (0.5%), to verify the effectiveness of each module in the proposed SDBDNet. The settings of the ablation experiments are as follows:

- To assess the function of the DropBlock module, we performed experiments without DropBlock (i.e. $r = 0$) while keeping other parameters unchanged (experiments on different values of r are presented in section 4.5 as part of the parameters experiments).
- To validate the role of weighted average, we set $\alpha \equiv 1$ and $\alpha \equiv 0$, respectively, while keeping other parameters unchanged. The former value corresponds to the condition when the sinogram before reconstruction is completely obtained by the sinogram-domain network and the latter corresponds to those without denoising (experiments on different values of β are presented in section 4.5 as part of the parameters experiments).
- To evaluate the effectiveness of joint training of the dual-domain networks, we turned off the back-propagation of the OS-SART layer, kept other parameters unchanged, and performed experiments. In this case, the training of the sinogram-domain network is not affected by the loss of the image domain but is wholly determined by the loss of the sinogram domain itself.

Table 4 presents the quantitative results of the ablation experiments with the low-dose conditions of $I_0 = 1 \times 10^4$ and $I_0 = 0.5 \times 10^4$, respectively. The PSNR and SSIM results demonstrate that each designed component makes a contribution to the denoising performance of the proposed SDBDNet. Specifically, the dual-domain joint training strategy, the DropBlock module in the sinogram domain, and the weighted average of the sinogram all helped to improve the denoising performance.

4.5. Parameters experiments

The proposed SDBDNet has three essential parameters: the dropout ratio r of the DropBlock layers, the scaling factor β of the weighted average, and the coefficient λ that balances the dual-domain losses. Our experiments on different datasets with different low-dose conditions show that the parameters for achieving good performance vary within a reasonable range. Specifically, in our experiments, we found that a dropout ratio of 0.2 or 0.3 for the DropBlock network and a scaling factor between 0.5 and 0.75 for the weighted average perform well. In fact, a too-large dropout ratio can undermine the denoising ability of the sinogram-domain network, while a too-small value may not effectively suppress blurring artifacts. With regard to the weighted average, a near half-to-half weight can better leverage the advantage of dual-domain denoising than a biased one. Moreover, considering that the weights are further normalized based on the maximum value, a scaling factor slightly greater than 0.5 is reasonable, which guides us in adjusting this parameter. Figure 10 shows the line charts of the average PSNR and SSIM results of the experiments with different β and r . We compared the results of the experiments on the LodoPaB dataset with the low-dose conditions of $I_0 = 0.5 \times 10^4$ (0.5%) and different β whose values were taken every 0.125 from 0 to 1. As shown in figure 10(a), as β increases, the average PSNR and SSIM show a trend of first increasing and then decreasing, and reach the highest values at $\beta = 0.5$ and 0.375 respectively. This reflects the impact of the weights on the final denoising performance and the importance of balancing the dual-domain denoising. In the experiments on the LoDoPaB dataset with the low-dose conditions of $I_0 = 1 \times 10^4$ (1%), we compared the average indicators of the results under different dropout ratios r whose values were taken every 0.1 from 0 to 0.5. As shown in figure 10(b), the average PSNR and SSIM also present an increasing and then decreasing trend with the raising of r , and both reach their maximums at $r = 0.2$. The comparison shows that the selection of the dropout ratios r has a notable impact on the denoising performance of SDBDNet. This is in line with the principle of denoising with the dropblock layers. In fact, a too-small r cannot

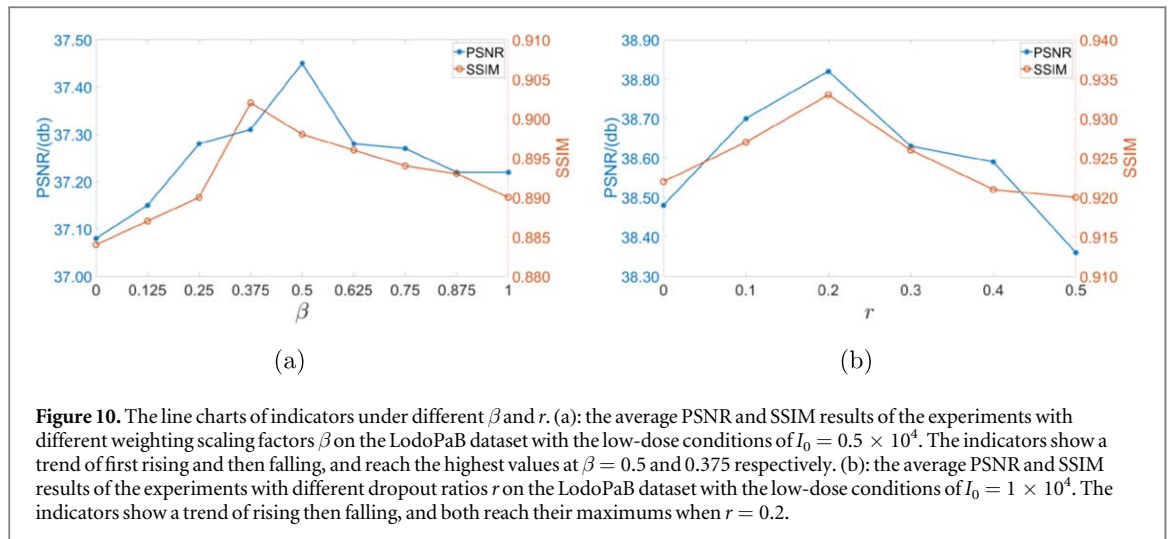


Figure 10. The line charts of indicators under different β and r . (a): the average PSNR and SSIM results of the experiments with different weighting scaling factors β on the LodoPaB dataset with the low-dose conditions of $I_0 = 0.5 \times 10^4$. The indicators show a trend of first rising and then falling, and reach the highest values at $\beta = 0.5$ and 0.375 respectively. (b): the average PSNR and SSIM results of the experiments with different dropout ratios r on the LodoPaB dataset with the low-dose conditions of $I_0 = 1 \times 10^4$. The indicators show a trend of rising then falling, and both reach their maximums when $r = 0.2$.

Table 5. comparison (psnr(db)/ssim) of different λ on the LodoPaB dataset with the low-dose conditions of $I_0 = 1 \times 10^4$ (1%) and $I_0 = 0.5 \times 10^4$ (0.5%).

	$I_0 = 1 \times 10^4$		$I_0 = 0.5 \times 10^4$	
λ	PSNR	SSIM	PSNR	SSIM
0.25	38.77	0.932	37.33	0.898
0.50	38.68	0.930	37.45	0.898
0.75	38.69	0.928	37.31	0.897
1.00	38.82	0.933	37.28	0.890
2.00	37.97	0.932	36.87	0.887

effectively suppress the blurring artifacts, while an over-large one will lead to excessive discarding of information, resulting in ineffective denoising.

We also performed experiments with different λ values to verify their impact on the denoising performance of SDBDNet on the LoDoPaB dataset with the low-dose conditions of $I_0 = 1 \times 10^4$ (1%) and $I_0 = 0.5 \times 10^4$ (0.5%). Table 5 presents the quantitative results. We found that, within a certain range, the difference of λ has little effect on the denoising results, but a too-large one could significantly reduce the denoising performance, which confirms the importance of maintaining dual-domain balance during network training.

5. Conclusion

We present SDBDNet, a self-supervised dual-domain deep learning method for LDCT denoising. SDBDNet conducts self-supervised denoising in both the sinogram and image domains and performs dual-domain joint learning to improve denoising performance. To generate paired training data with uncorrelated noise, SDBDNet splits sinograms into two parts and restores them with interpolation and learning-based correction. In addition, to achieve noise-adapted denoising and control the denoising strength in the sinogram domain, SDBDNet incorporates Dropblock layers and utilizes a weighted average of the treated and noisy sinograms. SDBDNet demonstrates high performance in LDCT denoising and outperforms popular non-supervised methods. In certain cases, it could even show advantages over supervised post-processing methods.

As a self-supervised method, SDBDNet offers greater application value than classical supervised approaches. The effective processing in the dual domain enables it to handle the details and structure of CT images well. In clinical work, its characteristic of not requiring paired or normal-dose data makes it more feasible. Physicians need only perform low-dose scans of patients to obtain the data for network training and inference, rather than potentially harmful normal-dose scans or repeated scans. For data belonging to different patients generated by the same machine with the same imaging parameters, they can be considered to approximately follow the same distribution and can be denoised using a trained or rapidly fine-tuned network. Although the training and inference time required by our SDBDNet are relatively long compared to image-domain methods, the fact that it does not necessitate repeated and complete training, and its inference speed remains fast, ensures its practicality.

Although SDBDNet shows good performance in self-supervised LDCT denoising, it still suffers from some limitations. Due to the use of traditional denoising networks in the image domain, SDBDNet, like other image domain methods, does not completely avoid blurring of the reconstructed images, which is reflected in our results. Second, embedding two reconstructions also makes SDBDNet require a lot of training and inference time beyond the image-domain methods. In addition, SDBDNet contains several non-learning hyperparameters, the adjustment during training will significantly increase the workload. In terms of the improvement for SDBDNet, one open challenge is how to generate paired data with independent noise while keeping the original noisy sinogram as complete as possible rather than splitting it in half. In addition, it is worth studying to reduce the number of hyperparameters while ensuring the denoising performance. Moreover, using fully-networked reconstruction to increase the efficiency of training and inference is also a meaningful research topic. We will focus on these issues in future research.

Acknowledgments

This work was supported by Beijing Natural Science Foundation (No. Z210003), National Natural Science Foundation of China (NSFC) (61971292) and China Scholarship Council (CSC) (No. 202307300001). The authors are also grateful to Beijing Higher Institution Engineering Research Center of Testing and Imaging for funding this research work.

Data availability statement

All data that support the findings of this study are included within the article (and any supplementary information files).

ORCID iDs

Ran An  <https://orcid.org/0000-0001-6558-1174>

Ke Chen  <https://orcid.org/0000-0002-6093-6623>

Hongwei Li  <https://orcid.org/0000-0002-9537-3808>

References

- Altekrüger F, Denker A, Hagemann P, Hertrich J, Maass P and Steidl G 2023 Patchnr: learning from very few images by patch normalizing flow regularization *Inverse Prob.* **39** 064006
- Andersen A H and Kak A C 1984 Simultaneous algebraic reconstruction technique (START): a superior implementation of the art algorithm *Ultrasound Imaging* **6** 81–94
- Batson J and Royer L 2019 Noise2self: blind denoising by self-supervision *Proceedings of the 36th International Conference on Machine Learning* 97 (PMLR) pp 524–33
- Buades A, Coll B and Morel J-M 2005 A non-local algorithm for image denoising *IEEE Computer Society Conf. on Computer Vision and Pattern Recognition (CVPR'05)* vol 2 (IEEE) pp 60–5
- Chen H, Zhang Y, Kalra M K, Lin F, Chen Y, Liao P, Zhou J and Wang G 2017 Low-dose CT with a residual encoder–decoder convolutional neural network *IEEE Trans. Med. Imaging* **36** 2524–35
- Creswell A, White T, Dumoulin V, Arulkumar K, Sengupta B and Bharath A A 2018 Generative adversarial networks: an overview *IEEE Signal Process. Mag.* **35** 53–65
- Croitoru F-A, Hondru V, Ionescu R T and Shah M 2023 Diffusion models in vision: a survey *IEEE Trans. Pattern Anal. Mach. Intell.* **45** 10850–69
- Dabov K, Foi A, Katkovnik V and Egiazarian K 2007 Image denoising by sparse 3D transform-domain collaborative filtering *IEEE Trans. Image Process.* **16** 2080–95
- de González A B, Mahesh M and Kim K-P 2010 Projected cancer risks from computed tomographic scans performed in the united states in 2007 *J. Vasc. Surg.* **51** 783
- Deng X, Zhao Y and Li H 2019 Projection data smoothing through noise-level weighted total variation regularization for low-dose computed tomography *J. X-ray Sci. Technol.* **27** 537–57
- Fang W, Wu D, Kim K, Kalra M K, Singh R, Li L and Li Q 2021 Iterative material decomposition for spectral CT using self-supervised noise2noise prior *Phys. Med. Biol.* **66** 155013
- Ge R, He Y, Xia C, Sun H, Zhang Y, Hu D, Chen S, Chen Y, Li S and Zhang D 2022 Ddpnet: a novel dual-domain parallel network for low-dose CT reconstruction *Int. Conf. on Medical Image Computing and Computer-assisted Intervention* (Springer) pp 748–57
- Ghiasi G, Lin T-Y and Le Q V 2018 Dropblock: a regularization method for convolutional networks *Advances in Neural Information Processing Systems 31 (NeurIPS 2018)* 31 (Curran Associates Inc.) 10750–60
- Gu J and Ye J C 2021 Adain-based tunable cyclegan for efficient unsupervised low-dose CT denoising *IEEE Trans. Comput. Imaging* **7** 73–85
- Guo S, Yan Z, Zhang K, Zuo W and Zhang L 2019 Toward convolutional blind denoising of real photographs *Proc. of the IEEE/CVF Conf. on Computer Vision and Pattern Recognition* pp 1712–22
- Hasan A M, Mohebbian M R, Wahid K A and Babyn P 2020 Hybrid-collaborative noise2noise denoiser for low-dose CT images *IEEE Trans. Radiat. Plasma Med. Sci.* **5** 235–44

- He Z, Zhang Y, Guan Y, Guan B, Niu S, Zhang Y, Chen Y and Liu Q 2022 Iterative reconstruction for low-dose CT using deep gradient priors of generative model *IEEE Trans. Radiat. Plasma Med. Sci.* **6** 741–54
- Hendriksen A A, Pelt D M and Batenburg K J 2020 Noise2inverse: self-supervised deep convolutional denoising for tomography *IEEE Trans. Comput. Imaging* **6** 1320–35
- Hore A and Ziou D 2010 Image quality metrics: PSNR versus SSIM *2010 20th Int. Conf. On Pattern Recognition (IEEE)* pp 2366–9
- Jin K H, McCann M T, Froustey E and Unser M 2017 Deep convolutional neural network for inverse problems in imaging *IEEE Trans. Image Process.* **26** 4509–22
- Kingma D P and Ba J 2014 Adam: a method for stochastic optimization arXiv:1412.6980
- Kobyzev I, Prince S J and Brubaker M A 2020 Normalizing flows: an introduction and review of current methods *IEEE Trans. Pattern Anal. Mach. Intell.* **43** 3964–79
- Krull A, Buchholz T-O and Jug F 2019 Noise2void-learning denoising from single noisy images *Proc. of the IEEE/CVF Conf. on Computer Vision and Pattern Recognition* pp 2129–37
- Kwon T and Ye J C 2021 Cycle-free cyclegan using invertible generator for unsupervised low-dose CT denoising *IEEE Trans. Comput. Imaging* **7** 1354–68
- Lehtinen J, Munkberg J, Hasselgren J, Laine S, Karras T, Aittala M and Aila T 2018 Noise2noise: learning image restoration without clean data *Proceedings of the 35th International Conference on Machine Learning* 80 (PMLR) 2965
- Leuschner J, Schmidt M, Baguer D O and Maass P 2021 Lodopab-CT, a benchmark dataset for low-dose computed tomography reconstruction *Sci. Data* **8** 109
- Li Z, Zhou S, Huang J, Yu L and Jin M 2020 Investigation of low-dose CT image denoising using unpaired deep learning methods *IEEE Trans. Radiat. Plasma Med. Sci.* **5** 224–34
- Liu X, Xie Y, Diao S, Tan S and Liang X 2023 A diffusion probabilistic prior for low-dose CT image denoising arXiv:2305.15887
- McCollough C 2016 TU-FG-207A-04: overview of the low dose CT grand challenge *Med. Phys.* **43** 3759–60
- Niu C, Li M, Fan F, Wu W, Guo X, Lyu Q and Wang G 2022 Noise suppression with similarity-based self-supervised deep learning *IEEE Trans. Med. Imaging* **42** 1590–602
- Niu C, Li M, Guo X and Wang G 2022 Self-supervised dual-domain network for low-dose CT denoising *Developments in X-ray Tomography XIV* vol 12242 (SPIE)
- Niu C et al 2021 Noise entangled gan for low-dose CT simulation arXiv:2102.09615
- Ronneberger O, Fischer P and Brox T 2015 U-net: convolutional networks for biomedical image segmentation *Medical Image Computing and Computer-assisted Intervention—MICCAI 2015: 18th Int. Conf., Munich, Germany, Proc., Part III* 18 9351 (Springer) pp 234–41
- Song Y and Ermon S 2020 Improved techniques for training score-based generative models *Adv. Neural Inf. Process. Syst.* **33** 12438–48
- Ulyanov D, Vedaldi A and Lempitsky V 2018 Deep image prior *Proc. of the IEEE Conf. on Computer Vision and Pattern Recognition* pp 9446–9454
- Wagner F et al 2023 On the benefit of dual-domain denoising in a self-supervised low-dose CT setting *2023 IEEE 20th Int. Symp. on Biomedical Imaging (ISBI) (IEEE)* pp 1–5
- Wang D, Fan F, Wu Z, Liu R, Wang F and Yu H 2023 Ctformer: convolution-free token2token dilated vision transformer for low-dose CT denoising *Phys. Med. Biol.* **68** 065012
- Wang J, Li T, Lu H and Liang Z 2006 Penalized weighted least-squares approach to sinogram noise reduction and image reconstruction for low-dose x-ray computed tomography *IEEE Trans. Med. Imaging* **25** 1272–83
- Won D, Jung E, An S, Chikontwe P and Park S H 2021 Self-supervised learning based CT denoising using pseudo-ct image pairs arXiv:2104.02326
- Wu D, Kim K and Li Q 2021 Low-dose ct reconstruction with noise2noise network and testing-time fine-tuning *Med. Phys.* **48** 7657–72
- Yaroslavsky L P 2012 *Digital Picture Processing: An Introduction* vol 9 (Springer Science & Business Media)
- You C et al 2019 Ct super-resolution gan constrained by the identical, residual, and cycle learning ensemble (GAN-circle) *IEEE Trans. Med. Imaging* **39** 188–203
- Yuan N, Zhou J and Qi J 2020 Half2half: deep neural network based CT image denoising without independent reference data *Phys. Med. Biol.* **65** 215020
- Zeng D, Huang J, Bian Z, Niu S, Zhang H, Feng Q, Liang Z and Ma J 2015 A simple low-dose x-ray CT simulation from high-dose scan *IEEE Trans. Nucl. Sci.* **62** 2226–33
- Zhang C, Chang S, Bai T and Chen X 2022 S2ms: self-supervised learning driven multi-spectral ct image enhancement *7th Int. Conf. on Image Formation in X-ray Computed Tomography* vol 12304 (SPIE) pp 473–9
- Zhang K, Zuo W, Chen Y, Meng D and Zhang L 2017 Beyond a gaussian denoiser: Residual learning of deep cnn for image denoising *IEEE Trans. Image Process.* **26** 3142–55
- Zhang K, Zuo W and Zhang L 2018 Ffdnet: toward a fast and flexible solution for cnn-based image denoising *IEEE Trans. Image Process.* **27** 4608–22
- Zhang Y, Hu D, Zhao Q, Quan G, Liu J, Liu Q, Zhang Y, Coatrieux G, Chen Y and Yu H 2021 Clear: comprehensive learning enabled adversarial reconstruction for subtle structure enhanced low-dose CT imaging *IEEE Trans. Med. Imaging* **40** 3089–101
- Zhang Z, Liang X, Zhao W and Xing L 2021 Noise2context: context-assisted learning 3D thin-layer for low-dose CT *Med. Phys.* **48** 5794–803
- Zhou B, Chen X, Xie H, Zhou S K, Duncan J S and Liu C 2022 Dudoufnet: dual-domain under-to-fully-complete progressive restoration network for simultaneous metal artifact reduction and low-dose CT reconstruction *IEEE Trans. Med. Imaging* **41** 3587–99
- Zhou L, Wang X, Hou M, Li P, Fu C, Ren Y, Shao T, Hu X, Sun J and Ye H 2022 Low-dose CT reconstruction by self-supervised learning in the projection domain arXiv:2203.06824
- Zhu Y, Zhao M, Zhao Y, Li H and Zhang P 2012 Noise reduction with low dose CT data based on a modified rof model *Opt. Express* **20** 17987–8004