frontiers in Signal Processing

ORIGINAL RESEARCH
published: 07 February 2022
doi: 10.3389/frsip.2021.812193

Low Dose CT Denoising by ResNet With Fused Attention Modules and Integrated Loss Functions

Luella Marcos[1], *Javad Alirezaie*[1]* *and Paul Babyn*[2]

[1]Department of Electrical and Computer Engineering, Ryerson University, Toronto, ON, Canada, [2]Department of Medical Imaging, Royal University Hospital and Saskatchewan Health Authority, University of Saskatchewan, Saskatoon, SK, Canada

X-ray computed tomography (CT) is a non-invasive medical diagnostic tool that has raised public concerns due to the associated health risks of radiation dose to patients. Reducing the radiation dose leads to noise artifacts, making the low-dose CT images unreliable for diagnosis. Hence, low-dose CT (LDCT) image reconstruction techniques have offered a new research area. In this study, a deep neural network is proposed, specifically a residual network (ResNet) using dilated convolution, batch normalization, and rectified linear unit (ReLU) layers with fused spatial- and channel-attention modules to enhance the quality of LDCT images. The network is optimized using the integration of per-pixel loss, perceptual loss *via* VGG16-net, and dissimilarity index loss. Through an ablation experiment, these functions show that they could effectively prevent edge oversmoothing, improve image texture, and preserve the structural details. Finally, comparative experiments showed that the qualitative and quantitative results of the proposed network outperform state-of-the-art denoising models such as block-matching 3D filtering (BM3D), Markovian-based patch generative adversarial network (patch-GAN), and dilated residual network with edge detection (DRL-E-MP).

Keywords: low-dose CT image, denoising, deep learning, residual learning, attention modules, per-pixel loss, perceptual loss, structural similarity index

OPEN ACCESS

Edited by:
Yunfeng Wu,
Xiamen University, China

Reviewed by:
Biao Wei,
Chongqing University, China
Yi Zhang,
Sichuan University, China

***Correspondence:**
Javad Alirezaie
javad@ryerson.ca

Specialty section:
This article was submitted to Biomedical Signal Processing, a section of the journal Frontiers in Signal Processing

Received: *09 November 2021*
Accepted: *13 December 2021*
Published: *07 February 2022*

Citation:
Marcos L, Alirezaie J and Babyn P (2022) Low Dose CT Denoising by ResNet With Fused Attention Modules and Integrated Loss Functions. Front. Sig. Proc. 1:812193. doi: 10.3389/frsip.2021.812193

INTRODUCTION

X-ray computed tomography (CT) is one of the most used diagnostic tools in medical imaging. It provides fine details of human internal structure noninvasively, which is ideal for detecting abnormalities in patients. However, the use of this image modality requires the use of X-rays to capture the region of interest. Exposure to such ionizing radiation can cause health risks including cancer (Z. Wang et al., 2020). Although some may argue that the effects of radiation from commercial CT scans is overstated, the dramatic expansion of the CT usage has already increased the global annual cumulative ionizing radiation dose by 34% (Tahmasebzadeh et al., 2021). Hence, researchers have been exploring effective ways to reduce radiation dose for medical imaging diagnosis without decreasing the accuracy of the image quality due to the added presence of noise.

Generally, radiation reduction is usually performed by controlling the X-ray current tube or by minimizing the X-ray photon count (Kulathilake et al., 2021). This process degrades the signal-to-noise ratio (SNR) of the X-ray signals, resulting in lower-quality CT images with noise artifacts, making clinical diagnosis less reliable. Various methods of radiation reduction have been introduced, which have already achieved improved results including sinogram domain filtering, iterative

reconstruction (IR), and image denoising using deep learning techniques, all of which aim to follow the "as low as reasonably achievable" (ALARA) principle (Yi and Babyn, 2018).

Projection domain filtering uses raw projection data before analytic CT image reconstruction. For noise removal, the noise present in the projection space should be well characterized (Wang et al., 2008). A recent study by Ma et al. (2021) proposed an attention deep residual dense convolutional neural network (CNN) with the intent of extracting noise features from the LDCT projection data in order to extract the clean sinogram for reconstruction. Although the fusion of the local and global feature information during this pre-processing of the sinogram data obtained pleasing results, acquiring raw sinogram data remains quite challenging from commercial CT scanners (Ma et al., 2021). Model-based iterative reconstruction (MBIR) techniques perform image reconstruction based on object projections. A continuous sequence of comparing an image assumption with the real time measured values for this method made it almost impossible for the early scanners to perform the method (Pickhardt et al., 2012). However, with the rapid advancement of computer technology, this technique can now be handled and can achieve higher image quality in terms of the image texture and spatial resolution (Hashimoto and Takamaeda-Yamazaki, 2021). Learned experts' assessment-based reconstruction network (LEARN) has been introduced, which utilizes the regularization and parameters used during the IR training process to effectively recover the images while trying to reduce the computational costs (Chen et al., 2018). A continued drawback is that the results are still susceptible to noise artifacts and the computational cost is high and similar to the sinogram domain filtering; there is also a limitation regarding the collection of projection data. Manifold and graph integrative convolutional (MAGIC) network simultaneously extracts pixel-level and topological features by using spatial and graph convolutions in an attempt to address the data limitation issue but still faces some potential issues regarding the optimization of the network design (Xia et al., 2021b).

For this study, CT image post-reconstruction is implemented using deep learning methods, which offers a more robust solution to overcome the issues regarding the mentioned iterative methods. Deep learning methods have been evolving throughout recent years and have been effectively providing reliable outcomes when applied in different fields especially in computer vision. These methods take advantage of the graphics processing unit (GPU) parallel computing in accelerating the training process when a network model contains deeper layers, which tends to have the vanishing gradient problem. Numerous state-of-the-art deep learning models have been developed in terms of reducing noise artifacts in LDCT images. Generally, the CT reconstruction process involves mapping features of normal-dose CT (NDCT) images with the low-dose images (LDCT), and this can be done through the denoising algorithms. Block-matching 3D (BM3D) is a transformation domain technique in which the same patches are stacked into 3D groups by block matching and transformed into wavelet domain during the reconstruction process (Dabov et al., 2007). Further, recent developments of generative adversarial networks (GANs) are also booming in the LDCT denoising research community due to the framework's ability to produce fine details of denoised images (Goodfellow et al., 2020). A GAN main framework typically consists of a generator that generates fake denoised images which will then be sent off to the discriminator. The discriminator gives a score on how fake denoised images compare with the NDCT images. This sequence repeats until the generated image becomes acceptable (L. Chen et al., 2020). Even though this framework certainly preserves the structural information of the images, problems like blurring remain noticeable. Sharpness-aware GAN (SAGAN) focused on addressing this problem of blurring effect and introduced an additional sharpness detection network for measuring the sharpness of the denoised image (Yi and Babyn, 2018). Moreover, boosting attention fusion GAN (BAFGAN) implements sub-modules that can include long-range dependencies of the LDCT images to produce higher-quality denoised images (Lyu et al., 2020). Similarly, U-Net-based discriminator in GAN framework (DU-GAN) simultaneously learns both local and global differences between the LDCT and NDCT images for a better regularization of the model (Huang et al., 2021). Although this framework can reliably provide exceptional outputs, the deep and complex architecture is also prone to instability due to the oscillating number of parameters during the training process. The local parameters for each sub-network of GAN must be trained as well as the parameters of the overall GAN during the training process, which is the main challenge with GAN architectures. As the accuracy of the discriminator increases, the performance of the generator gets worse during the training process. The unbalanced performance of the discriminator and the generator can cause vanishing gradient, making the whole system unstable (Arjovsky and Bottou, 2017).

A simpler but more stable denoising structure is the use of residual network (ResNet), in which skip connections between pre- and post-convolutional layers during the denoising process are implemented (He et al., 2016). The structure of a residual network provides decreased computational costs than GANs without deteriorating the quality of the denoised images. A residual encoder–decoder CNN (RED-CNN) demonstrates the effectiveness of using symmetric convolution and deconvolutional network using skip connections in denoising LDCT images at high computational speed (Chen et al., 2017a; Chen et al., 2017b). A parameter-dependent framework (PDF)-based RED-CNN network has also been introduced, which is trained simultaneously via two multilayer perceptrons (MLPs) that are used for modulating the feature maps of CT reconstruction process (Xia et al., 2021a). A ResNet merged with U-Net is able to learn both local and global image features, avoiding the vanishing gradient system, which is similar to the objective of DU-GAN but has a very comprehensive architecture while achieving the same results (Liu et al., 2021). The feasibility of a residual neural network was also explored by applying the concept of transfer learning for LDCT image denoising especially when an unknown noise level is present (Zhong et al., 2020). In addition, dilated residual learning with an edge detection layer (DRL-E-MP), composed of a Sobel kernel, integrated the advantages of having dilated convolutions

instead of the standard convolution and symmetric shortcut connections for conserving the data features as well as capturing the structural details at the image boundaries better (Gholizadeh-Ansari et al., 2019). Further, a similar network uses a dilated residual learning with perceptual loss and structural dissimilarity (DRLPS), in which the focus is to take into consideration the structural detail in low contrast regions (Ataei et al., 2020a).

Inspired by DRL-E-MP, DRLPS, and BAFGAN denoising models (Gholizadeh-Ansari et al., 2019; Ataei et al., 2020a; Lyu et al., 2020), fused attention modules in dilated residual learning network (FAM-DRL) is introduced. This proposed network applies the concept of the attention modules from BAFGAN. Since BAFGAN has a complex architecture and faced instability issues, the proposed denoiser utilizes dilated convolutional layers and skip connections for faster network training, better stability, and more effective fusion of the feature attention modules. In this experiment, FAM-DRL would be optimized using the combination of perceptual loss via VGG-16 Net for the prevention of edge oversmoothing, structural dissimilarity loss (DSSIM) for texture enhancement, and per-pixel loss for the symmetry between NDCT and LDCT images (Kulathilake et al., 2021). The main contribution of this paper is the unique architecture of the proposed denoising network which achieves the following:

1) protection of edges from blurring,
2) enhancement of image textures, and
3) preservation of structural details of the CT images.

The remainder of this paper is organized as follows: *Network Architecture* provides full detail of the components used for the proposed network; *Experiments* discusses the data, training details and environment, and the evaluation method for the experiment. The *Results* section presents the quantitative and visual results, followed by the *Discussion* section where analytic observations are documented. Finally, the *Conclusion* summarizes the overall findings of this study.

NETWORK ARCHITECTURE

In this section the proposed network containing the fused attention modules for the fusion of spatial- and channel-wise features of the images is presented.

Proposed Dilated Residual Network

Shown in **Figure 1**, the proposed denoiser network is constructed using 3×3 dilated convolution layers with a dilation rate of 2, batch normalization (BN), and ReLU layers in order to extract the shallow features. Further, the number of filters used for each convolutional layers follows the standard setting of 64 (Zhong et al., 2020). For this process, 512×512 LDCT images, x, are used as an input. More details about the datasets are discussed in *Experiments* section of this paper. Next is the generation of the multi-dimensional deep features in the cascaded boosting module groups (BMG). For this experiment, three BMG blocks are implemented. In each BMG, a stack of $n \epsilon \{1, \ldots, N\}$ boosting attention fusion blocks (BAFB) contain the fusion of the spatial and channel attention modules as shown in **Figure 2**, which will be further discussed in *Fused Attention Modules*. Lastly, deconvolution + BN + ReLU make up the reconstruction layers as represented by the three post-convolutional layers after the BMG modules in **Figure 1**. To prevent the vanishing gradient problem, symmetric skip connection (SSC) between the pre-and post-convolutional blocks are applied. To test the accuracy of the overall network, peak signal-to-noise ratio (PSNR) and structural similarity index metrics (SSIM) are used for comparing the structural information of the NDCT-LDCT image pairs.

Fused Attention Modules

The long-range dependencies of the CT image can be obtained by passing the input through several convolutional layers. A simple Conv + BN + ReLU operation cannot simply achieve the high- and low-frequency information of the feature map present during the pre-convolutional process. Hence, as demonstrated inside the BAFBs in **Figure 1**, the integration of a spatial and channel-attention modules are implemented, which is shown in **Figure 2A**. The fusion of these boosting attention modules captures the long-range dependencies of the image during the feature extraction process. The structure of the attention modules is based on the boosting modules used in BAFGAN (Lyu et al., 2020). Without additional supervision, the fused attention mechanism allows the network to focus on the most relevant features. Hence, avoiding the use of similar feature maps instead highlights the primary features that are useful for LDCT denoising tasks (Sinha and Dolz, 2021).

Spatial Attention Module

On the one hand, the spatial attention module (SAM), $f_{\mathrm{SAM}}(\cdot)$, in **Figure 2B** uses the feature maps obtained from the third convolutional block, $f_{CR1}(x)$, of the network as an input. The process can be represented as follows:

$$F_{up} = f_{SAM1}^{up}\left(f_{CR1(x)}\right) \oplus f_{CR1}(x) \tag{1}$$

Further, it uses SoftMax activation function also known as the normalized exponential function for a smoother normalization in different dimensions, making each component to be in the interval [0, 1]. This helps in incorporating the prior assumptions based on the topological spatial-wise in the structure of the image. For this spatial network, the assumption is that the feature vectors would be dependent on each other in a spatially smooth consistent way (Miladinovi`c et al., 2021). The main purpose is to improve the performance of FAM-DRL with the additional spatial dependency layers, shown in **Figure 2B**.

Channel Attention Module

On the other hand, the channel attention module (CAM), $f_{\mathrm{CAM}}(\cdot)$, also uses the same input as SAM, but this module also captures the channel-wise features instead of capturing the long-range dependencies only. The channel attention module pipeline is demonstrated in **Figure 2C**, which can also be represented as follows:

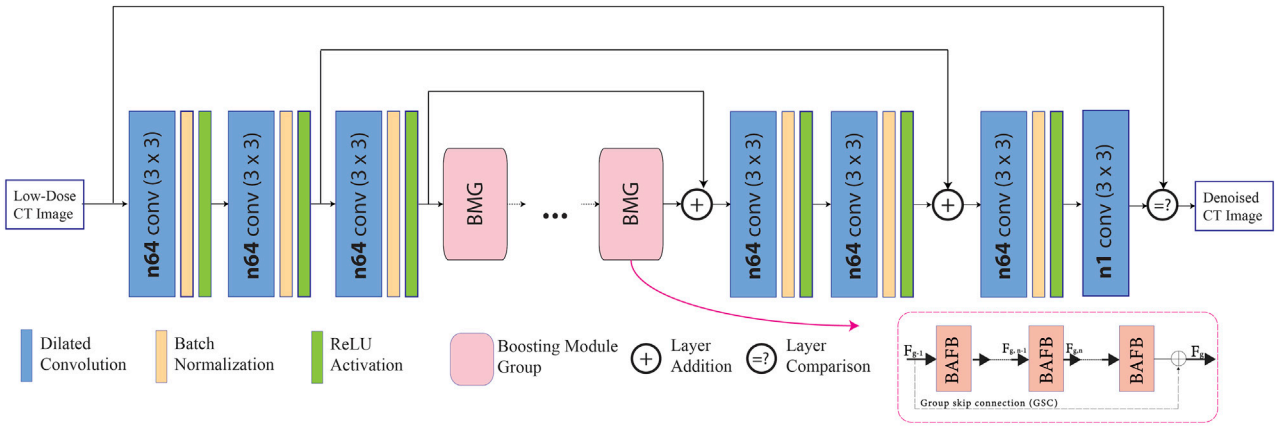

FIGURE 1 | Proposed fused attention modules in dilated residual network, in which the attention modules are located inside the boosting attention fused block (BAFB).

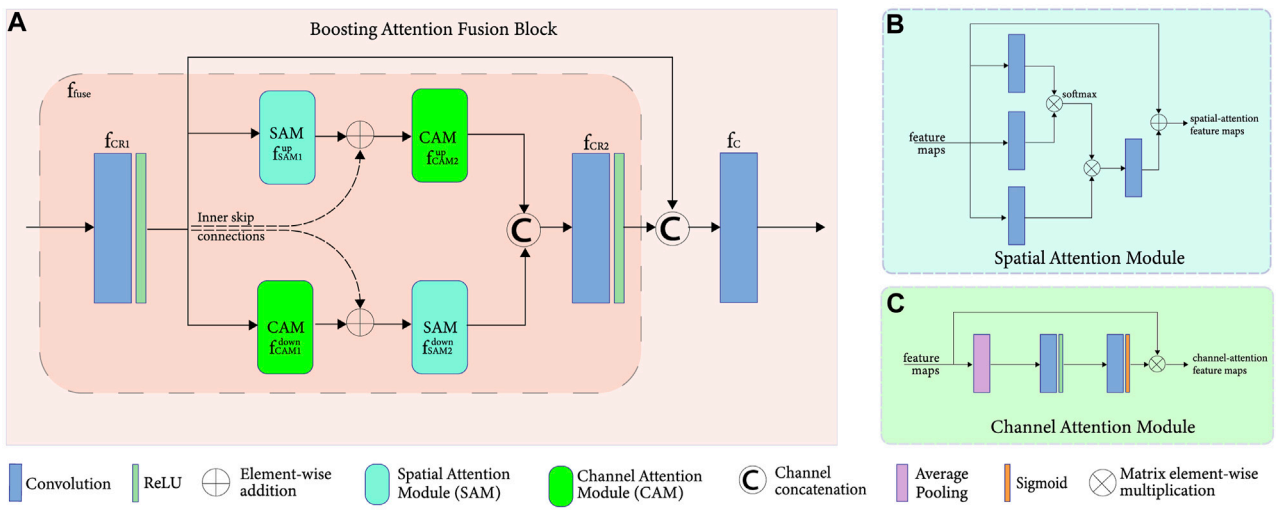

FIGURE 2 | (A) Boosting attention fusion blocks (BAFB) composed of **(B)** spatial attention and **(C)** channel attention modules.

$$F_{down} = f_{CAM1}^{down}\left(f_{CR1}(x)\right) \oplus f_{CR1}(x) \tag{2}$$

This module uses average pooling, which permits a small amount of invariance in the image and could extract more features than normal max pooling. This enhances the features from all the channels, increasing feature discriminability for preserving structural details of the image. A sigmoid activation function or the logistic function is used to also capture nonlinearities, which allows the network to learn more complex structures in the data.

Overall Attention Module

In order for the spatial and channel-wise characteristics to complement each other, a fusion, $f_{fused}(\cdot)$, between the two is applied as well as implementing inner skip connections. Mathematically,

$$f_{fuse} = F_{up} \copyright F_{down} \copyright f_{CAM2}^{up}\left(F_{up}\right) \copyright f_{SAM2}^{down}\left(F_{down}\right) \tag{3}$$

where $\oplus$ denotes element-wise addition and $\copyright$ represents channel concatenation in this case. At the end of this module, the new generated features are fed into a convolutional layer, f_c, producing the spatial-channel attention features.

Loss Functions

For this research, the combination of three loss functions 1) mean-squared error (MSE), 2) perceptual loss, and 3) structural similarity index is proposed for the optimization of the overall network.

Per Pixel Loss

Mean squared error (MSE), considered as a per-pixel loss function, is one of the most common accuracy measurements that calculates the difference between the LDCT, x_i and NDCT, y_i images. Then, all the absolute errors between pixels are added:

$$L_{MSE} = \frac{1}{N}\sum_{i=1}^{N}\|y_i - x_i^2\| \tag{4}$$

The application of MSE can cause oversmoothing problem along the edges of CT images during the training process as observed in a CycleGAN and FFDNet denoising models (Zhang et al., 2018; Gu and Ye, 2021).

Perceptual Loss

In order to address blurring issue, the proposed model also utilizes the perceptual loss calculated from using the VGG16-pretrained network (Simonyan and Zisserman, 2015). Unlike MSE, perceptual loss takes high level features into consideration in order to more accurately correspond to the human visual system. This is due to its ability of learning the features more accurately as proven in DRL-E-MP and cascaded CNN (Gholizadeh-Ansari, Alirezaie, and Babyn 2019; Ataei et al., 2020b). This perceptual loss utilizes the feature maps, ϕ_i, that are extracted from the last convolutional layer in blocks i = 1, 2, 3, 4 of the VGG16Net with size $h_i \times w_i \times d_i$, which can be expressed as follows:

$$L_{PL} = \sum_{i=1}^{N}\frac{1}{h_i w_i d_i}\left\|\phi_i(x) - \phi_i(y)\right\|^2 \tag{5}$$

Structural Similarity Index Metrics

Finally, structural similarity index metrics (SSIM) have the ability to compare the structural information of the image such as the texture, contrast, luminance, and the compression (Kulathilake et al., 2021). The SSIM between the LDCT and NDCT can be calculated as follows:

$$SSIM(x, y) = \frac{\left(2\mu_x\mu_y + c_1\right)\left(2\sigma_{xy} + c_2\right)}{\left(\mu_x^2 + \mu_y^2 + c_1\right)\left(\sigma_x^2 + \sigma_y^2 + c_2\right)} \tag{6}$$

where μ, σ, and σ_{xy} stand for the mean, sample standard deviation, and sample covariance, respectively. However, this cannot be applied directly to the network as a loss function since the objective of this expression is to maximize the output value close to 1 and would provide higher values as the loss. Therefore, structural dissimilarity (DSSIM) expressed in **Eq. 7** is implemented which is the SSIM equivalent as a kernel loss function.

$$L_{SSIM} = \frac{1 - SSIM(x, y)}{2} \tag{7}$$

Overall Objective Function

The overall objective function for the proposed network can be represented as follows:

$$L\left(\hat{Y}, Y\right) = \frac{\gamma_1 L_{MSE}\left(\hat{Y}, Y\right) + \gamma_2 L_{PL}\left(\hat{Y}, Y\right) + \gamma_3 L_{SSIM}\left(\hat{Y}, Y\right)}{\sum_{i=1}^{n}\gamma_i} \tag{8}$$

where γ_1, γ_2, γ_3 are the sum-to-one weights for the three loss components and $(\hat{Y}, Y)$ is the LDCT and ground-truth image pair. Each of the weights is determined during the training process, where the maximum value of the losses after each epoch is used for updating the values of the weights. The loss function that obtained the greatest loss would receive a higher scale than the other functions.

EXPERIMENTS

Dataset

For this research, five different datasets are used: NDCT–LDCT image pairs of a deceased Piglet and Phantom Thoracic datasets by Yi and Babyn and three more datasets from the *Cancer* Imaging Archive Mayo Clinic are provided (Yi and Babyn, 2018; McCollough et al., 2021). The Piglet and Thoracic datasets are simulated and evaluated by expert radiologists including the co-author, Dr. Paul Babyn. The clinical datasets from the Mayo clinic (Abdomen, Chest, and Head) are gathered from the American Association of Physicists in Medicine (AAPM) grand challenge database, which contains normal-dose and simulated low-dose images and have more realistic noise assumptions. Head and abdomen datasets are provided at 25% of the routine dose, and chest cases are provided at 10% of the routine dose (McCollough et al., 2021). The specification for each dataset is summarized in **Table 1**. The standard partition, 70–30%, of the training and testing is applied. In addition, the training dataset of size 512 × 512 is subdivided into 32 × 32 overlapping patches to increase the number of training samples and minimize the computational load of the network.

Training Environment

The training operation of the model especially with the parameters in the BMGs are the same with implementation done with BAFGAN (Lyu et al., 2020). The proposed network for this research was trained for 200 epochs with a batch size of 4 and using the ADAM optimizer with a learning rate of 0.0002, β_1 = 0.01, and β_2 = 0.999. The implementation of this model was done with Tensorflow-Keras API on Windows operating system with Intel® Core™i7 CPU @2.80 GHz processor and NVIDIA GeForce GTX 1080 graphics card.

Evaluation

To further strengthen the validity of the proposed structure of the proposed loss functions, three modifications of the network model have been done: FAM-DRL with only MSE, FAM-DRL with only perceptual loss, FAM-DRL with only DSSIM, and finally, FAM-DRL with integration of the three loss functions as the proposed model. The results are also compared with the results from the implementation of DRL-E-MP (Gholizadeh-Ansari et al., 2019), modified BM3D (Dabov et al., 2007; Makinen et al., 2020), and self-attentive spectral normalized Markovian patch-GAN or modified patch-GAN (Bera and

TABLE 1 | NDCT-LDCT image dataset specifications.

Dataset	Number of pairs	Scanner specs
Piglet	146	100 KVp, 0.625 mm thickness, 300 mAs (NDCT), 15 mAs (LDCT)
Thoracic	90	120 KVp, 0.75 mm thickness, 480 mAs (NDCT), 60 mAs (LDCT)
Abdomen	209	100 KVp, 5 mm thickness, 502 mAs (NDCT), 498 mAs (LDCT)
Chest	278	120 KVp, 5 mm thickness, 712 mAs (NDCT), 690 mAs (LDCT)
Head	40	120 KVp, 5 mm thickness (unknown X-ray current tube info)

TABLE 2 | Summary of the average PSNR and SSIM of the denoising algorithms for Piglet, Thoracic, Abdomen, Chest, and Head datasets.

Models	Piglet		Thoracic		Abdomen		Chest		Head	
	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM
Modified BM3D	40.23	0.8444	28.73	0.4461	38.12	0.8602	30.24	0.4461	39.82	0.5677
Patch GAN	39.23	0.7524	30.37	0.5435	38.32	0.8746	31.28	0.5435	40.86	0.6217
DRL-E-MP	41.66	0.9529	27.31	0.4181	37.64	0.8578	31.33	0.6636	38.64	0.4893
FAM-DRL (MSE)	41.64	0.9313	30.36	0.5459	39.84	0.8834	33.23	0.5544	41.51	0.5272
FAM-DRL (PL)	41.15	0.9272	30.42	0.5888	39.97	0.8862	33.84	0.5623	41.57	0.5798
FAM-DRL (SSIM)	40.48	0.9687	29.24	0.6342	38.44	0.8992	32.03	0.6674	39.93	0.6723
FAM-DRL (proposed)	**42.93**	**0.9765**	**31.07**	**0.6388**	**40.33**	**0.9102**	**34.26**	**0.6971**	**42.64**	**0.6870**

The bold values highlights the highest PSRN/SSIM value for each column.

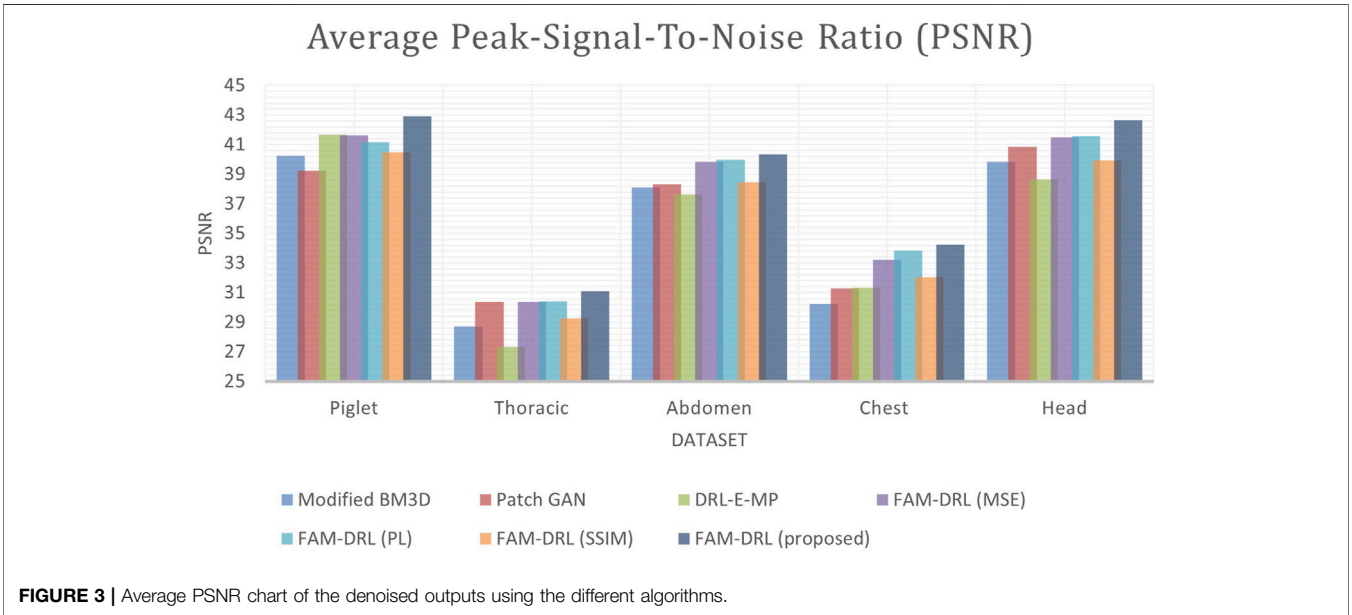

FIGURE 3 | Average PSNR chart of the denoised outputs using the different algorithms.

Biswas, 2020) models. The quantitative results are mainly based on the PSNR and SSIM.

RESULTS

Quantitative Results

This section provides the quantitative results of the variation of the models as well as the different algorithms: 1) modified BM3D, 2) patch-GAN, 3) DRL-E-MP, 4) FAM-DRL with MSE, 5) FAM-DRL with perceptual loss (PL), 6) FAM-DRL with SSIM, and 7) the proposed FAM-DRL with MSE + PL + MSE. **Table 2** summarizes the PSNR and SSIM obtained, while **Figures 3** and **4** show separate charts for the trend of models in terms of PSNR and SSIM, respectively.

Each model of the models is run using the five datasets in order to obtain the average PSNR and SSIM. For the Piglet dataset, the average PSNR of the models ranges from 39 to 42, while the average ranges from 0.7 to 0.9 as shown in **Table 2**. In terms of PSNR, it shows in **Figure 3** that the proposed FAM-DRL has gained a slightly higher improved PSNR (42.93) compared to the other models for Piglet dataset. This

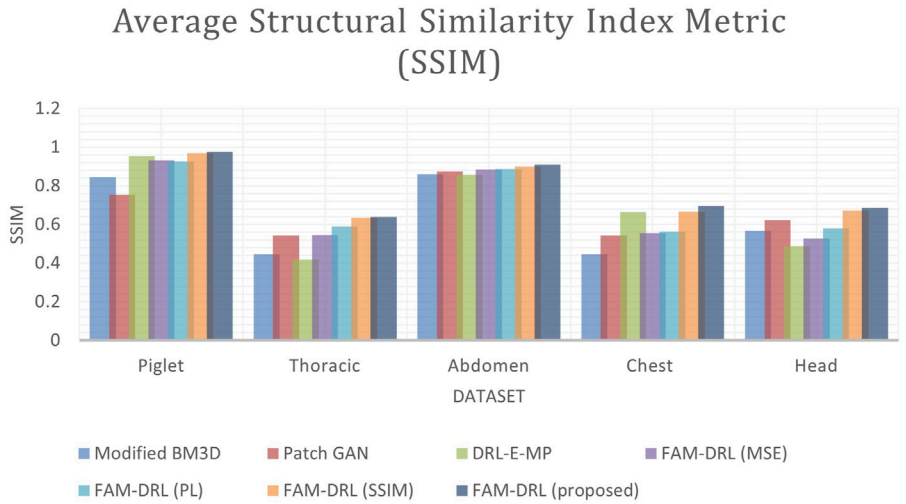

FIGURE 4 | Average SSIM of the outputs using the different algorithms.

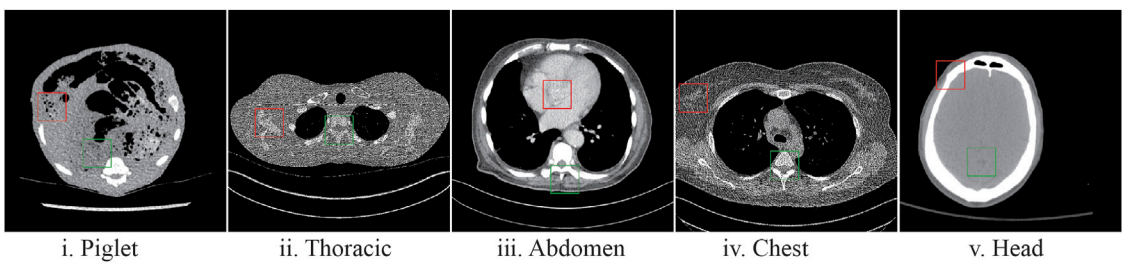

FIGURE 5 | Sample LDCT image slices for each dataset with the marked regions (**Figures 6–10**), where the difference between algorithms and ground-truth images are more noticeable.

pattern can also be observed with the PSNR results when using the other datasets. Observing the PSNR trend in **Figure 3**, FAM-DRL with only perceptual loss (FAM-DRL-PL), FAM-DRL with only MSE loss (FAM-DRL-MSE), and DRL-E-MP are comparatively close and can be considered second best after the proposed model. The modified BM3D and patch-GAN show huge difference when compared to the PSNR of the proposed network.

Looking at the SSIM trend in **Figure 4**, the difference between the average SSIM of the models is slightly smaller using the different datasets. Despite these small gaps between the SSIM of the models, the proposed model with the integration of the objective functions still ranks first when it comes to the highest SSIM. Moreover, FAM-DRL with only SSIM loss function ranks second as expected since the use of SSIM as loss function aims to minimize the distinction of the structural information between the NDCT and LDCT image pairs. As for the other models, it shows that there is no clear pattern of which model comes next after the proposed model and the model with only SSIM kernel function. This discrepancy is due to the variation of the structural information of the different datasets.

Visual Results

In **Figure 5** sample results are displayed utilizing the first slice of each dataset. The marked regions in **Figure 5** correspond to structural details of the image where the differences between the algorithms are pronounced.

The marked regions as shown in each dataset slice image in **Figure 5** are highlighted in **Figures 6i–10i** along with the visual results of the algorithms, **Figures 6–10ii–viii**. Investigating the visual results of BM3D, there is an obvious oversmoothing problem that can be observed in **Figures 6**, **8**, **10ii** as well as apparent checkbox artifacts in **Figures 7**, **8**, **9ii**. Markovnian patch-GAN and FAM-DRL (MSE) show slightly better visual results than BM3D but still display similar problems. This is due to the fact that these algorithms only use MSE as loss function, which is well known for causing oversmoothing along the edges. In comparison to visual results of FAM-DRL with only perceptual loss in **Figures 6iii–10iii**, FAM-DRL (SSIM) shows evident artifacts in **Figures 6–10iv**. Despite the apparent artifacts, FAM-DRL (SSIM) is able to preserve the textural details of the images. The combination of these three objective functions embedded in the proposed network presents well-structured denoised images closer to the NDCT or ground-truth images,

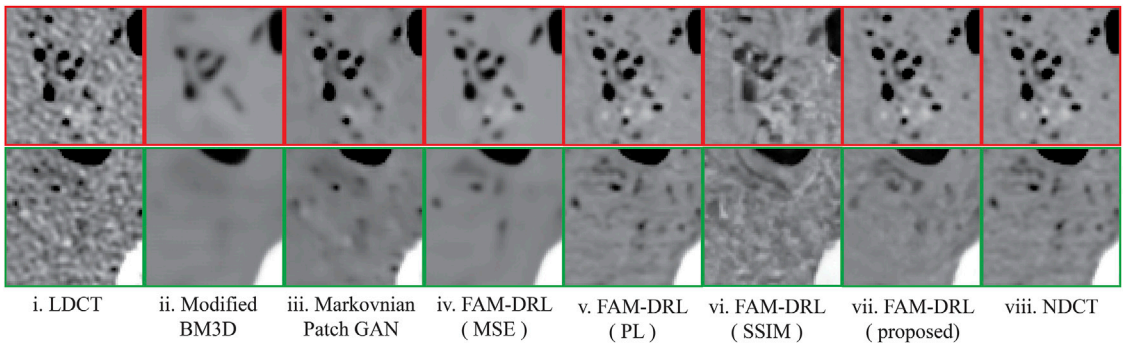

FIGURE 6 | Sample results of the highlighted sections from **Figure 5i** from first slice of the LDCT deceased piglet dataset using the different algorithms.

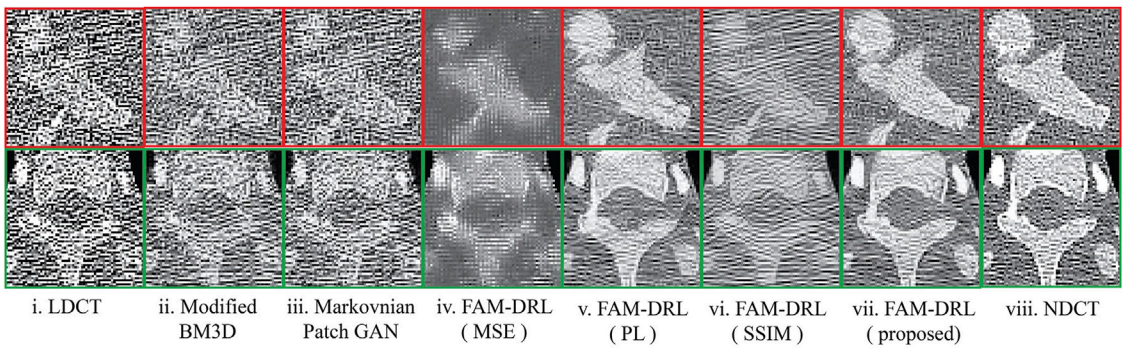

FIGURE 7 | Sample results of the highlighted sections from **Figure 5ii** from first slice of the LDCT phantom thoracic dataset using the different algorithms.

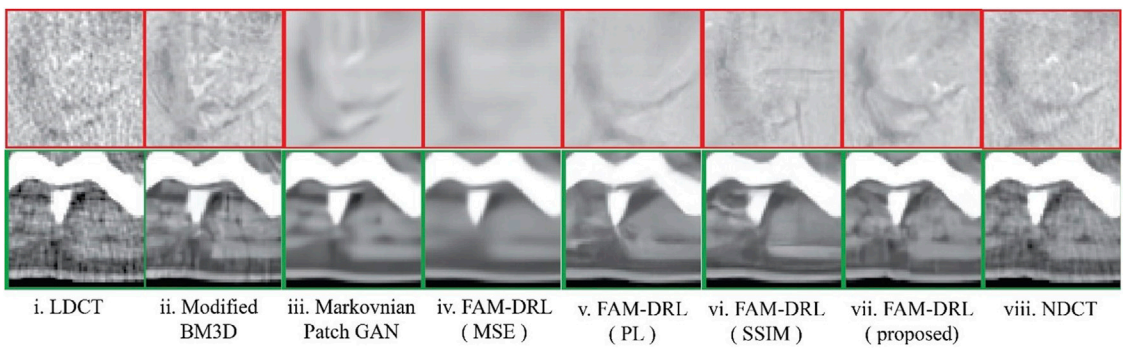

FIGURE 8 | Sample results of the highlighted sections from **Figure 5iii** from first slice of the LDCT abdomen (patient L058) dataset using different algorithms.

as demonstrated in **Figures 6vii–10vii** and **Figures 6viii–10viii**, respectively.

DISCUSSION

The overall results show that the proposed FAM-DRL with the integration of the three loss functions outperforms the benchmark models as well as variations of the model itself. While FAM-DRL with only perceptual loss obtained higher PSNR compared with the other two variations of the proposed model (FAM-DRL with MSE, FAM-DRL with SSIM), FAM-DRL with only SSIM gained higher values in terms of SSIM as demonstrated in **Figure 4**. For the visual results, oversmoothing along edges is noticeable when only MSE was applied to the network; enhancement of the perceptual quality is visible but introduced some abnormalities when perceptual loss is used, and image texture is more enhanced when SSIM is applied.

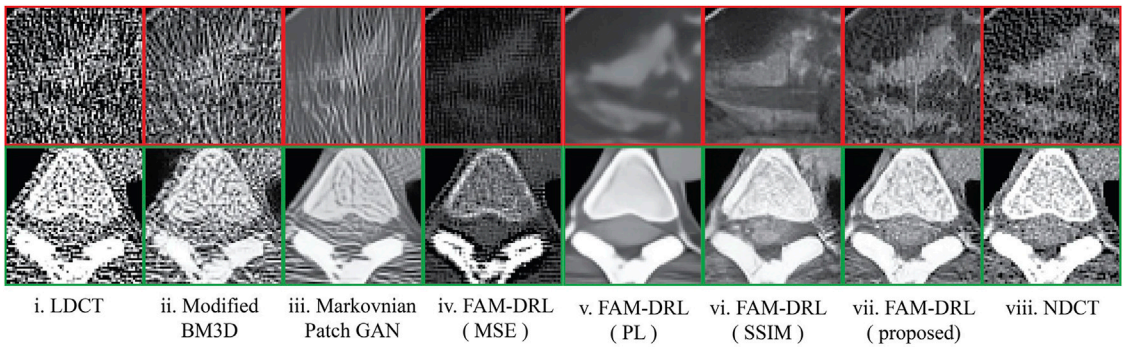

FIGURE 9 | Sample results of the highlighted sections from **Figure 5iv** from first slice of the LDCT Chest (patient C002) dataset using different algorithms.

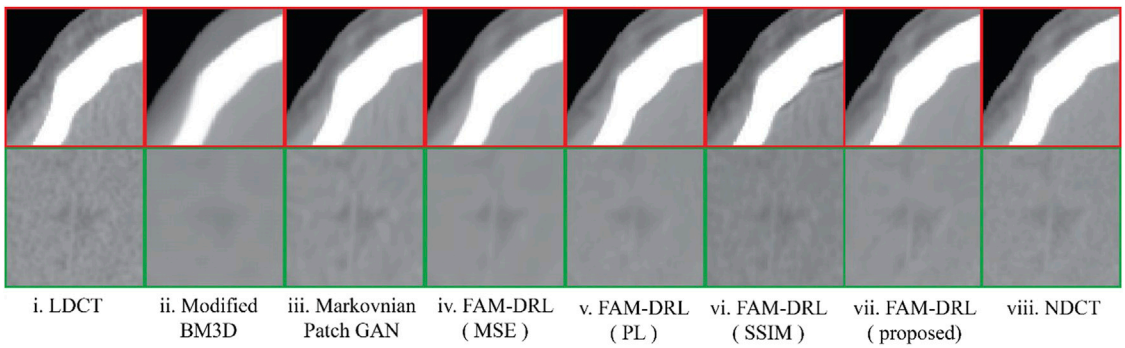

FIGURE 10 | Sample results of the highlighted sections from **Figure 5v** from first slice of the LDCT Head (patient N012) dataset using different algorithms.

Despite the drawbacks displayed by each loss function, the output of the combination of the three in the network complement their limitations individually. Hence, the overall proposed model shows promising results when compared to the state-of-the-art models.

The modified BM3D and patch-GAN acquired the lowest PSNR and SSIM values, summarized in **Table 2**, which are slightly lower than the FAM-DRL with MSE loss function only. These models implemented the use of MSE loss function. Although the outputs from MSE are acceptable quantitatively, this does not guarantee having appealing visual results since this loss function typically causes blurring effects. The regions shown in **Figures 6–10** correspond to structural details of the image where the differences between the algorithms are most pronounced as marked in **Figure 5**. The visual results of the models that utilized MSE as loss functions are shown in **Figures 6–10**ii–iii. Based on these results, blurring effects and noise artifacts stand out when compared to the variations of the proposed models.

According to **Table 2**, the PSNR/SSIM for DRL-E-MP is really close to the results obtained for FAM-DRL with perceptual loss only and with the proposed FAM-DRL with the combination of the objective functions. This is due to the fact that both models used the same perceptual loss functions derived from the same blocks in VGG16-Net. This can also be observed in the PSNR and SSIM trend in **Figures 3** and **4**, respectively, not only with the quantitative results but also with the visual results as demonstrated in **Figures 6–10**, in which the images show the specific regions selected for each dataset. The models that use perceptual loss display more natural and perceptually appealing results. Even though the use of perceptual loss seems effective enough, it can introduce some anomaly due to regularization and hyper-parameter tuning since the perceptual loss applied for this experiment used the pre-trained VGG16-Net. For example, there is an apparent generation of fracture in **Figure 7**v that could indicate a remodelled bone. Blurring effects can also be seen in **Figures 8** and **9**, which contain fine details of the images.

That being said, when perceptual loss was combined with the other loss functions, the proposed model obtained the highest PSNR and SSIM compared with the other models for all the datasets used. The overall visual result of the proposed model contains the textural details close to the ground-truth image while also maintaining high SNR and avoiding the oversmoothing problem from applying MSE. Therefore, this study can be deemed as successful for it meets the expected results experimentally.

Although the improvements shown in this paper are due to the fused attention modules, which the benchmark models do not have, another comparison could be done for further research which focuses on the effectiveness of using the attention modules

by testing the accuracy of the architecture with and without the attention modules.

Moreover, comparison metrics such as contrast-to-noise ratio (CNR) and noise power spectrum (NPS) for measuring the intensity difference at low-contrast regions and texture quality of the LDCT and NDCT paired images could be used for comparative studies (Brombal et al., 2019). However, CNR does not capture visibility dependence of the image structure on the detail size, which PSNR can measure. Moreover, using NPS for measuring the accuracy would also require the background of the images to be removed since it is highly dependent on the various characteristic image parameters (Dolly et al., 2016). This includes the size and number of the region of interest, which should be constant within in the images; otherwise it would cause statistical fluctuations. Therefore, a Phantom-based study is desired for future work to examine the image quality of denoised LDCT images in terms of CNR and NPS measures.

In addition, almost all existing deep learning-based approaches, like the proposed network, usually require LDCT and NDCT paired training datasets. However, there is no guarantee to have paired LDCT and NDCT images readily available. The acquisition of the paired datasets for post-reconstruction of CT images can be from multiple scans, like the datasets provided by the Mayo Clinic, or from data simulations for producing matches from unpaired data, like the datasets provided by Yi and Babyn. For the network model to be trained without LDCT-NDCT image pairs, unsupervised learning method is recommended.

CONCLUSION

In this experiment, it was shown that creating feature maps by implementing the fusion of spatial- and channel-attention modules can enhance the SNR of the images. The use of dilated convolutions and skip connections, main components of the proposed model, also provided efficiency for the possible increased computational costs of the model that can be commonly seen in GAN-based denoising models. In addition, we also demonstrated the individual contribution and limitations of each objective functions used in the network such as perceptual loss for enhancement of the perceptual visual results and SSIM kernel loss function for image enhancement.

DATA AVAILABILITY STATEMENT

The original contributions presented in the study are included in the article/Supplementary Material; further inquiries can be directed to the corresponding author.

AUTHOR CONTRIBUTIONS

All authors contributed to the study conception and design. The first draft of the manuscript was written by LM, and all authors commented and provided feedback to the versions of the manuscript for improvement. All authors have read and approved the final manuscript.

FUNDING

This research has been supported by the NSERC Discovery grant RGPIN-2020–04441 that is awarded to JA.

ACKNOWLEDGMENTS

The authors would like to thank Cynthia McCollough, the Mayo Clinic, and the American Association of Physicists in Medicine for making the CT data available for the study.

REFERENCES

Arjovsky, M., and Bottou, L. (2017). Towards Principled Methods for Training Generative Adversarial Networks. *Stat* 1050, 1–17.

Ataei, S., Alirezaie, J., and Babyn, P. (2020b). Cascaded Convolutional Neural Networks with Perceptual Loss for Low Dose CT Denoising. *Int. Jt. Conf. Neural Netw. (IJCNN)*, 1–5.

Ataei, S., Alirezaie, J., and Babyn, P. (2020a). "Low Dose CT Denoising Using Dilated Residual Learning with Perceptual Loss and Structural Dissimilarity," in Middle East Conference on Biomedical Engineering, 2020-October (MECBME), 8–12. doi:10.1109/MECBME47393.2020.9265165

Bera, S., and Biswas, P. K. (2021). Noise Conscious Training of Non Local Neural Network Powered by Self Attentive Spectral Normalized Markovian Patch GAN for Low Dose CT Denoising. *IEEE Trans. Med. Imaging* 40, 3663–3673. doi:10.1109/TMI.2021.3094525

Brombal, L., Arfelli, F., Delogu, P., Donato, S., Mettivier, G., Michielsen, K., et al. (2019). Image Quality Comparison between a Phase-Contrast Synchrotron Radiation Breast CT and a Clinical Breast CT: a Phantom Based Study. *Sci. Rep.* 9 (1), 1–12. doi:10.1038/s41598-019-54131-z

Chen, H., Zhang, Y., Chen, Y., Zhang, J., Zhang, W., Sun, H., et al. (2018). LEARN: Learned Experts' Assessment-Based Reconstruction Network for Sparse-Data CT. *IEEE Trans. Med. Imaging* 37 (6), 1333–1347. doi:10.1109/TMI.2018.2805692

Chen, H., Zhang, Y., Kalra, M. K., Lin, F., Chen, Y., Liao, P., et al. (2017a). Low-Dose CT with a Residual Encoder-Decoder Convolutional Neural Network. *IEEE Trans. Med. Imaging* 36 (12), 2524–2535. doi:10.1109/tmi.2017.2715284

Chen, H., Zhang, Y., Zhang, W., Liao, P., Li, K., Zhou, J., et al. (2017b). aLow-dose CT via Convolutional Neural Network. *Biomed. Opt. Express* 8 (2), 679. doi:10.1364/boe.8.000679

Chen, L., Zheng, L., Lian, M., and Luo, S. (2020). A C-gan Denoising Algorithm in Projection Domain for Micro-CT. *MCB Mol. Cell Biomech.* 17 (2), 85–92. doi:10.32604/mcb.2019.07386

Dabov, K., Foi, A., Katkovnik, V., and Egiazarian, K. (2007). Image Denoising by Sparse 3-D Transform-Domain Collaborative Filtering. *IEEE Trans. Image Process.* 16 (8), 2080–2095. doi:10.1109/tip.2007.901238

Dolly, S., Chen, H.-C., Anastasio, M., Mutic, S., and Li, H. (2016). Practical Considerations for Noise Power Spectra Estimation for Clinical CT Scanners. *J. Appl. Clin. Med. Phys.* 17 (3), 392–407. doi:10.1120/jacmp.v17i3.5841

Gholizadeh-Ansari, M., Alirezaie, J., and Babyn, P. (2019). Deep Learning for Low-Dose CT Denoising Using Perceptual Loss and Edge Detection Layer. *J. Digit Imaging* 33, 504–515. doi:10.1007/s10278-019-00274-4

Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., et al. (2020). Generative Adversarial Networks. *Commun. ACM* 63 (11), 139–144. doi:10.1145/3422622

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. doi:10.1109/TCI.2021.3050266

Hashimoto, N., and Takamaeda-Yamazaki, S. (2021). An FPGA-Based Fully Pipelined Bilateral Grid for Real-Time Image Denoising. *An FPGA-Based Fully Pipelined Bilateral Grid for Real-Time Image Denoising* 1, 167–173. doi:10.1109/fpl53798.2021.00035

He, K., Zhang, X., Ren, S., and Sun, J. (2016). Identity Mappings in Deep Residual Networks. *Computer Vis. – ECCVLNCS* 9908 (1), 630–645. doi:10.1007/978-3-319-46493-0_38

Huang, Z., Zhang, J., Zhang, Y., and Shan, H. (2021). DU-GAN: Generative Adversarial Networks with Dual-Domain U-Net Based Discriminators for Low-Dose CT Denoising. *IEEE Trans. Instrum. Meas.*, 1. doi:10.1109/TIM.2021.3128703

Kulathilake, K. A. S. H., Abdullah, N. A., Sabri, A. Q. M., and Lai, K. W. (2021), A Review on Deep Learning Approaches for Low-Dose Computed Tomography Restoration, *Complex Intell. Syst.* 1-33. doi:10.1007/s40747-021-00405-x

Liu, J., Kang, Y., Qiang, J., Wang, Y., Hu, D., and Chen, Y. (2021). Low-dose CT Imaging via Cascaded ResUnet with Spectrum Loss. *Methods* S1046-2023 (21), 00131–00136. doi:10.1016/j.ymeth.2021.05.005;

Lyu, Q., Guo, M., and Ma, M. (2020). Boosting Attention Fusion Generative Adversarial Network for Image Denoising. *Neural Comput. Applic* 33, 4847–4833. doi:10.1007/s00521-020-05284-w

Ma, Y.-J., Ren, Y., Feng, P., He, P., Guo, X.-D., and Wei, B. (2021). Sinogram Denoising via Attention Residual Dense Convolutional Neural Network for Low-Dose Computed Tomography. *Nucl. Sci. Tech.* 32 (4), 1–14. doi:10.1007/s41365-021-00874-2

Makinen, Y., Azzari, L., and Foi, A. (2020). Collaborative Filtering of Correlated Noise: Exact Transform-Domain Variance for Improved Shrinkage and Patch Matching. *IEEE Trans. Image Process.* 29, 8339–8354. doi:10.1109/TIP.2020.3014721

McCollough, C. H., Chen, B., Holmes, D., III Duan, X., Yu, Z., Yu, L., et al. (2021). Data from Low Dose CT Image and Projection Data [Data Set]. *The Cancer Imaging Archive*. doi:10.7937/9npb-2637

Miladinovi`c, Đ. de., Stani´, A., Bauer, S., Schmidhuber, J., and Buhmann, J. M. (2021). "Spatial Dependency Networks: Neural Layers for Improved Generative Image Modeling," in International Conference on Learning Representations, 1–15. Available at: https://openreview.net/forum?id=I4c4K9vBNny.

Pickhardt, P. J., Lubner, M. G., Kim, D. H., Tang, J., Ruma, J. A., del Rio, A. M., et al. (2012). Abdominal CT with Model-Based Iterative Reconstruction (MBIR): Initial Results of a Prospective Trial Comparing Ultralow-Dose with Standard-Dose Imaging. *Am. J. Roentgenology* 199 (6), 1266–1274. doi:10.2214/AJR.12.9382.Abdominal

Simonyan, K., and Zisserman, A. (2015). "Very Deep Convolutional Networks for Large-Scale Image Recognition," in 3rd International Conference on Learning Representations, ICLR 2015 - Conference Track Proceedings, 1–14. arXiv preprint arXiv:1409.1556.

Sinha, A., and Dolz, J. (2021). Multi-Scale Self-Guided Attention for Medical Image Segmentation. *IEEE J. Biomed. Health Inform.* 25 (1), 121–130. doi:10.1109/JBHI.2020.2986926

Tahmasebzadeh, A., Paydar, R., Soltani-kermanshahi, M., Maziar, A., and Reiazi, R. (2021). Lifetime Attributable Cancer Risk Related to Prevalent CT Scan Procedures in Pediatric Medical Imaging Centers. *Int. J. Radiat. Biol.* 97 (9), 1282–1288. doi:10.1080/09553002.2021.1931527

Wang, J., Lu, H., Liang, Z., Eremina, D., Zhang, G., Wang, S., et al. (2008). An Experimental Study on the Noise Properties of X-ray CT Sinogram Data in Radon Space. *Phys. Med. Biol.* 53 (12), 3327–3341. doi:10.1088/0031-9155/53/12/018.An

Wang, Z., Lv, M.-Y., and Huang, Y.-X. (2020). Effects of Low-Dose X-Ray on Cell Growth, Membrane Permeability, DNA Damage and Gene Transfer Efficiency. *Dose-Response* 18 (4), 155932582096261–11. doi:10.1177/1559325820962615

Xia, W., Lu, Z., Huang, Y., Liu, Y., Chen, H., Zhou, J., et al. (2021a). CT Reconstruction with PDF: Parameter-dependent Framework for Data from Multiple Geometries and Dose Levels. *IEEE Trans. Med. Imaging* 40 (11), 3065–3076. doi:10.1109/TMI.2021.3085839

Xia, W., Lu, Z., Huang, Y., Shi, Z., Liu, Y., Chen, H., et al. (2021b). MAGIC: Manifold and Graph Integrative Convolutional Network for Low-Dose CT Reconstruction. *IEEE Trans. Med. Imaging* 40 (12), 3459–3472. doi:10.1109/tmi.2021.3088344

Yi, X., and Babyn, P. (2018). Sharpness-Aware Low-Dose CT Denoising Using Conditional Generative Adversarial Network. *J. Digit Imaging* 31, 655–669. doi:10.1007/s10278-018-0056-0

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 (9), 4608–4622. doi:10.1109/TIP.2018.2839891

Zhong, A., Li, B., Luo, N., Xu, Y., Zhou, L., and Zhen, X. (2020). Image Restoration for Low-Dose CT via Transfer Learning and Residual Network. *IEEE Access* 8, 112078–112091. doi:10.1109/ACCESS.2020.3002534

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors, and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.