



Deep Learning for Low-Dose CT Denoising Using Perceptual Loss and Edge Detection Layer

Maryam Gholizadeh-Ansari¹ · Javad Alirezaie^{1,2} · Paul Babyn³

Published online: 12 September 2019
© Society for Imaging Informatics in Medicine 2019

Abstract

Low-dose CT denoising is a challenging task that has been studied by many researchers. Some studies have used deep neural networks to improve the quality of low-dose CT images and achieved fruitful results. In this paper, we propose a deep neural network that uses dilated convolutions with different dilation rates instead of standard convolution helping to capture more contextual information in fewer layers. Also, we have employed residual learning by creating shortcut connections to transmit image information from the early layers to later ones. To further improve the performance of the network, we have introduced a non-trainable edge detection layer that extracts edges in horizontal, vertical, and diagonal directions. Finally, we demonstrate that optimizing the network by a combination of mean-square error loss and perceptual loss preserves many structural details in the CT image. This objective function does not suffer from over smoothing and blurring effects causing by per-pixel loss and grid-like artifacts resulting from perceptual loss. The experiments show that each modification to the network improves the outcome while changing the complexity of the network, minimally.

Keywords Low-dose CT image · Dilated convolution · Deep neural network · Noise removal · Perceptual loss · Edge detection

Introduction

Computed tomography (CT) is an accurate and non-invasive method to detect abnormalities in the internal parts of the body for instance tumors, bone fractures, and vascular diseases. In the past decades, it has been widely used by clinicians to diagnose and monitor conditions such as cancer, lung nodules, and internal injuries.

Since the CT images are produced by omitting X-ray beams at the body, there has been a growing concern about the risk of radiation. The amount of exposure during one session of CT scan is much higher than a conventional X-ray. For example, the radiation that a patient receives

in a chest X-ray radiography is equal to 10 days of background radiation [1]. Background radiation is the amount of radiation that a person gets from cosmic and natural resources in daily life. During a chest CT scan, the radiation exposure is equal to two years of background radiation [1]. Therefore, the radiation risk is much higher in computed tomography especially for those who had multiple CT scans. While radiation affects all the age groups, children are more vulnerable than adults because of their developing body and the longer lifespan which means more CT scans may be needed. Research has found that children who have cumulative doses from multiple head scans are up to three times more at risk of diseases such as leukemia and brain tumors [2].

Considering the advantages of CT scans diagnosis, it is critical to find a solution to the radiation problem. One approach to decreasing the radiation risk is to use lower doses of X-ray current. However, the produced CT images are not as clear and detailed as normal-dose CT images. Therefore, they will not be reliable for diagnosis. This need has made noise removal from low-dose CT an active research field.

In recent years, many types of research have been conducted to enhance the quality of the reconstructed

✉ Javad Alirezaie
javad@ryerson.ca

¹ Department of Electrical and Computer Engineering, Ryerson University, 350 Victoria Street, Toronto, ON M5B2K3, Canada

² Department of Systems Design Engineering, University of Waterloo, Waterloo, ON, Canada

³ Department of Medical Imaging, University of Saskatchewan and Saskatoon Health Region, Saskatoon, SK, Canada

CT images. Researches have followed three paths to remove noise from low-dose CT images: processing the raw data obtained from sinogram (projection space denoising), iterative reconstruction methods, and processing reconstructed CT image (image space denoising) [3].

In projection space denoising, the noise removal algorithm is applied to the CT sinogram data obtained from low-dose X-ray beams. Sinogram data, also called projection data or raw data, is a 2-D signal that represents the sum of the attenuation coefficients for a beam passing through the body. The noise distribution of low-dose CT image in the projection space can be well-characterized [4, 5] which makes the noise removal task simple. Some researchers have applied traditional noise removal techniques on this data such as bilateral filtering before image reconstruction. [6, 7]. These methods incorporate system physics and photon statistics to reduce both noise and artifacts. However, it makes the algorithm vendor dependent. These methods also need access to sinogram data which is not available for many commercial CT scanners. Finally, these techniques should be implemented on the scanner reconstruction system that increases the cost of denoising [3].

Iterative reconstruction methods are another group to improve the quality of low-dose CT images [8, 9]. In these methods, the data is transformed into the image domain and projection space multiple times to optimize the objective function. In the first step, a CT image is reconstructed using the projection data and then it is transformed back to the projection space. In each iteration, the generated projection data from the reconstructed CT image is compared with the actual data from the scanner and gets corrected. The process stops when the convergence criteria are met. These methods take into account the system model geometry, photon counting statistics, as well as x-ray beam spectrum, and they usually outperform the projection space denoising methods. They are capable of removing artifacts and providing good spatial resolution. However, similar to the previous group, they need access to the projection data, are vendor dependent, and should be implemented on the reconstruction system of the scanner. Moreover, the process is slow, and the computational cost of multiple iterations is high [3].

Opposite to the previous methods, image space denoising algorithms do not need the projection data. They work directly on the reconstructed CT images and are generally fast, independent of the scanner vendor and can be easily integrated with the workflow. Many of the proposed algorithms in this category are adopted from natural image processing. KSVD [10] is a dictionary learning algorithm based on sparse representation and dictionary learning. It is used for tasks such as image denoising, feature extraction, and image compression. In some studies, KSVD

is employed to improve the quality of low-dose CT scans [11, 12]. Non-local means [13] is another algorithm initially proposed for image denoising that has also been used by researchers for low-dose CT image enhancement [14]. The method calculates a weighted mean of the pixels in the image based on their similarity to the target pixel. The state-of-the-art block matching 3D (BM3D) [15] is also proposed for dealing with natural image noise. It is similar to the non-local means but works in a transform domain like wavelet or discrete Fourier transform. The first step of BM3D is to group patches of the image that have similar local structure and then stack them and form a 3-dimensional array. After transforming the data, a filter is applied to remove the noise. This method has been followed in some studies to perform low-dose CT noise removal [16, 17].

In recent years, many advances have been made in the image processing field by using deep learning (DL). The high computational capacity of GPUs in combination with techniques such as batch normalization [18] and residual learning [19] has made training deep networks possible. Some of the proposed networks have outperformed traditional methods in challenging tasks such as image segmentation, image recognition, and image enhancement. Medical imaging has also benefitted from this advancement. One of the first networks to reduce the noise of the low-dose CT image was proposed by Chen et al. [20]. It was inspired by a network designed for image super-resolution with three convolutional layers [21]. Convolutional auto-encoders have been used in [22] and [23] while the later also takes advantage of residual learning. All of the mentioned networks offer an end to end solution for low-dose noise removal. They receive a low-dose CT image as an input and predict the normal-dose CT image as the output. However, Kang et al. firstly finds the wavelet coefficients for low-dose and normal-dose CT images [24]. Then these wavelet coefficients are given to a 24-layer convolutional network as data (input) and labels (output). The inverse wavelet transform should be performed on the output results to find the normal-dose CT image.

Generative adversarial networks (GAN) are a group of deep neural networks that were first introduced by Good-Fellow [25]. GAN has two sub-networks, a generative network (G) and a discriminative network (D) that are trained simultaneously. The discriminative network is responsible for detecting real data from fake data while the generative network tries to create fake data as close as possible to the real data and fool the discriminator. Generative adversarial networks have attracted much interest, and researchers have applied it to different fields such as text to image synthesis [26], image super-resolution [27] and video generation [28]. GAN has also been used

to remove noise from low-dose CT images [29–31], where the generative network receives the low-dose CT images. It generates normal-dose images that the discriminative network cannot distinguish them from real normal-dose images.

In this paper, we have proposed a deep neural network to remove noise from low-dose CT images. Figure 1 displays this network. One approach to achieving higher performance in deep learning is to increase the number of layers which has become possible after introducing residual learning [19] and batch normalization [18]. However, more layers essentially mean more weights and higher computational cost. In this research, we have looked for methods that enhance the efficiency of the network without adding to its complexity. For this purpose, our network employs batch normalization, residual learning, and dilated convolution to perform denoising. We have also introduced an edge detection layer that improves the results with little increase in the number of training weights. The edge detection layer extracts edges in four directions and helps to enhance the performance. Finally, we have shown that optimizing the mean-square error as the loss function do not capture all the texture details of the CT image. For this purpose, we have used a combination of perceptual loss and mean-square error (MSE) as an objective function that significantly improves the visual quality of the output and keeps the structural details. The perceptual loss is used in GAN to generate fake images that are visually close to the target image by comparing the feature maps of two images. Yang et al. [32] have used the perceptual loss for CT image denoising but they compared the predicted image and the ground truth with one group of feature maps. In this study, feature maps have been extracted from four blocks of pre-trained VGG-16 [33] and used as a comparison tool in conjunction with the mean-square error.

Methods

Low-Dose CT Simulation

One of the challenges in applying machine learning techniques to the medical domain is the shortage of training samples. A neural network learns the probability distribution of the data from all the samples that it sees during the training process. If there are not sufficient samples to train the network for all conditions, the prediction will not be accurate. To train a network for low-dose denoising, we need normal-dose and low-dose pairs obtained in similar conditions. Obtaining such a dataset was not easy. For this reason, we have generated a low-dose dataset from normal-dose CT images to be used for training besides two other datasets that we had.

According to the literature, the dominant noise of a low-dose CT image in the projection domain has Poisson distribution [4, 5]. Therefore, to simulate a low-dose CT image, we have added Poisson noise to the sinogram data of the normal-dose image. The following steps show this procedure [34, 35]:

1. Compute the Hounsfield unit numbers of the normal-dose CT image HU_{nd} from its pixel values, using the Eq. 1 (if the CT image has padding, it should be removed, first),

$$HU_{nd} = \text{PixelValue} \times \text{Slope} + \text{Intercept} \quad (1)$$

Slope and Intercept values can be found from DICOM header.

2. Compute the linear attenuation coefficients μ_{nd} based on water attenuation μ_{water} [34],

$$\mu_{nd} = \frac{\mu_{\text{water}}}{1000} HU_{nd} + \mu_{\text{water}} \quad (2)$$

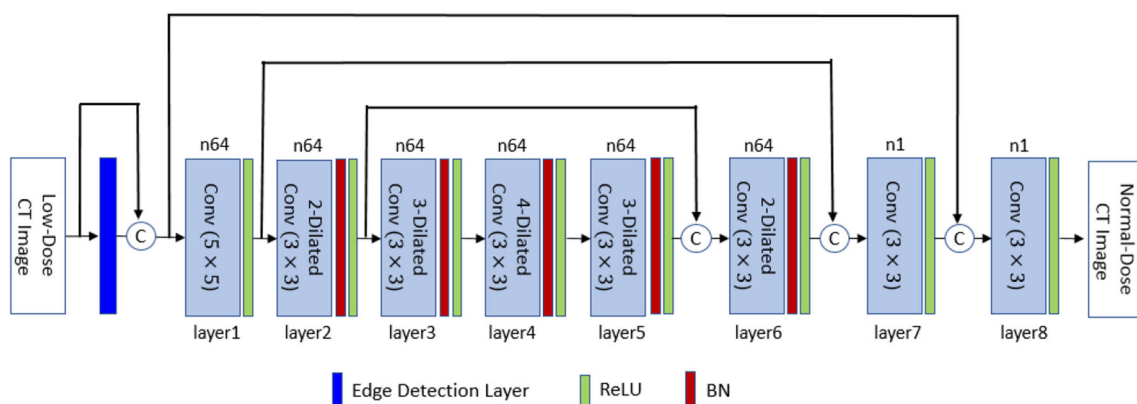


Fig. 1 Architecture of the proposed network. BN stands for batch normalization, i-Dilated Conv represents convolution operator with dilation rate i ($i=2,3,4$), and the activation function is the rectified linear unit (ReLU). Operator © performs concatenation

3. Obtain the projection data for normal-dose image ρ_{nd} by applying radon transform on linear attenuation coefficients [34, 35]. To eliminate the size factor, this should be multiplied by the voxel size.
4. Compute the normal-dose transmission data T_{nd} [35],

$$T_{nd} = \exp(-\rho_{nd})$$

5. Generate the low-dose transmission T_{ld} by injecting Poisson noise [35],

$$T_{ld} = \text{Poisson}(I_{ld}^0 T_{nd}) \tag{3}$$

here, I_{ld}^0 is simulated low-dose scan incident flux.

6. Calculate the low-dose projection data ρ_{ld} ,

$$\rho_{ld} = \ln\left(\frac{I_{ld}^0}{T_{ld}}\right)$$

7. Find the projection of the added Poisson noise [34],

$$\rho_{\text{noise}} = \rho_{nd} - \rho_{ld}$$

8. Compute the linear attenuation of the low-dose CT image μ_{ld} [34],

$$\mu_{ld} = \mu_{nd} + \text{iradon}\left(\frac{\rho_{\text{noise}}}{\text{voxel}}\right)$$

where, iradon represents the inverse Radon transform.

9. Finally apply the inverse of Eq. 2 to find the Hounsfield unit numbers for the low-dose CT image [34]. Figure 2d demonstrates a normal-dose image and the simulated low-dose images with different incident flux I^0 .

Dilated Convolution

Dilated convolution was introduced to deep learning in 2015 [36, 37] to increase the receptive field faster. Receptive field (RF) is the region of the input image that is used to calculate the output value. Larger receptive field means that more contextual information from the input image is captured. The classical methods to grow the receptive field are employing pooling layers, larger filters, and more layers in the network. A pooling layer performs downsampling and is a powerful technique to increase the receptive field. Although it is widely used in classification tasks, adoption

of a pooling layer is not recommended in denoising or super-resolution tasks. Downsampling with a pooling layer may lead to the loss of some useful details that cannot be recovered completely by upsampling methods such as transposed convolution [38]. Utilizing larger filters or more layers increases the number of weights drastically, meaning larger memory resources will be needed. Dilated convolution, also called atrous convolution, makes it possible to increase the receptive field with just a fraction of weights. One-dimensional dilated convolution is defined as

$$y[i] = \sum_{k=1}^f x[i + r.k]w[k] \tag{4}$$

where $x[i]$ and $y[i]$ are the input and the output of the dilated convolution. w represents the weight vector of the filter with length f , and r is the dilation rate.

Receptive field of the layer L (RF_L) with filter size $f \times f$ and dilation rate of r can be computed from Eq. 5 [39].

$$RF_L = RF_{L-1} + (f - 1)r \tag{5}$$

Equation 6 computes the number of weights needed for an N-layer convolutional network with a filter size $f \times f$.

$$\text{number of weights} = n \times f^2 \times c + n^2 \times f^2 \times (N-2) + n \times f^2 \times c \tag{6}$$

here, n is the number of filters in each layer and c is the number of channels. For simplification, we assume all the layers have n filters and the number of the channels in the input and output images are same. Table 1 compares the number of weights and layers needed to achieve receptive field equal to 13 in different cases. According to this table, using the dilation rate $r = 3$ lets to achieve the desired receptive field in only 4 layers when 3×3 filters are used.

To better understand the capability of dilated convolution, Wang et al. replaced the standard convolutions in [40] with dilated convolutions with $r = 2$ and achieved comparable performance in only 10 layers instead of 17 layers [39].

In this research, we have used an 8-layer dilated convolutional network to remove noise from low-dose CT

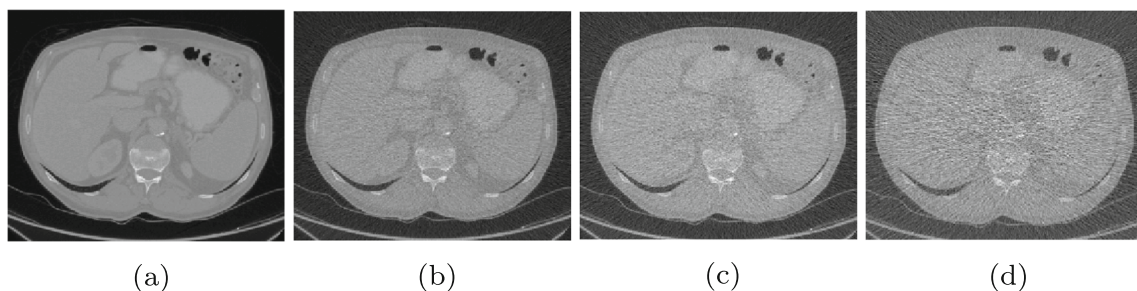


Fig. 2 Simulation of a low-dose CT image from Lung dataset. **a** Normal-dose CT image, simulated low-dose image with **b** $I_{ld}^0 = 1 \times 10^4$, **c** $I_{ld}^0 = 5 \times 10^3$, and **d** $I_{ld}^0 = 2^3$

Table 1 Number of training weights to obtain $RF = 13$ with different filter sizes. The number of filters in each layer is $n = 64$. The dilation rate in equal to r

| Filter size | 3×3 $r = 1$ | 5×5 $r = 1$ | 7×7 $r = 1$ | 3×3 $r = 3$ |
|---------------------------------------|-------------------------|-------------------------|-------------------------|-------------------------|
| Number of layers needed for $RF = 13$ | 6 | 5 | 4 | 4 |
| Number of weights | 148,608 | 310,400 | 407,680 | 748,800 |

images. The proposed network was inspired from a study by Zhang et al. [41], and the dilation rates are 1, 2, 3, 4, 3, 2, 1, and 1 for layers 1 to 8.

Residual Learning

One approach to improving the performance of a network is stacking more layers; nevertheless, researchers observed that networks with more layers do not always perform better. Opposite to the expectations, it has been seen that in a deeper network even the training loss grows. This degradation problem implies that the optimization of a deep network is not as easy as a shallow one. He et al. [19] proposed a residual learning framework to solve this problem by adding an identity mapping between the input and the output of a group of layers. From then, researchers have investigated many different combinations of adding shortcuts between different layers and achieved interesting results [38, 42].

In this study, we have exploited the residual learning to improve the performance of the network. Our experiments showed that adding symmetric shortcuts between the bottom and top layers boosts the performance. As Fig. 1 displays, the input image and the output of layers 2 and 3 are concatenated with the output of layers 7, 6 and 5, respectively. These connections pass the details of the image to higher layers, as feature maps in the first layers contain a lot of information from the input.

Edge Detection Layer

In image processing, edge detection refers to techniques that find the boundaries of objects in an image. Many of these methods search for discontinuities in the image brightness that are generally the result of the existence of an edge. Researchers have developed some advanced algorithms to extract edges from the image. In this study, we have adopted a simple edge detection technique to enhance the outcome of our network. Sobel edge detection operator [43] computes the 2-D gradient of the image intensity and emphasizes the regions with high spatial frequency by convolving the image with a 3×3 filter. The proposed edge detection layer is a convolutional layer that has four Sobel kernels as the non-trainable filters. Figure 3 shows the output of this layer for a low-dose CT image. These outputs are concatenated with the input image and given to the network. Our experiments confirm that the edge detection layer improves the performance of the network.

Objective Function

Mean-square error (MSE) is widely used as an objective function in low-level image processing tasks such as image denoising or image enhancement. MSE computes the difference of intensity between the pixels of output and the ground truth images. It is also used in many of the proposed algorithms for low-dose CT denoising. We started our research by optimizing MSE, but we noticed that the results do not express all the details of a CT image, though peak signal to noise (PSNR) is relatively high. This problem has been seen in image super-resolution tasks too [44]; however, it is more pronounced in CT images as the image is seen in a DICOM viewer with different grey-level mappings (windowing). Windowing helps to highlight the appearance of different structures and make a diagnosis. Our experiments showed that MSE loss generates blur images that do not include all the details, especially in the textures.

Johnson et al. demonstrated that using a perceptual loss achieves visually appealing results [44]. To compute the perceptual loss, the ground truth image and the predicted

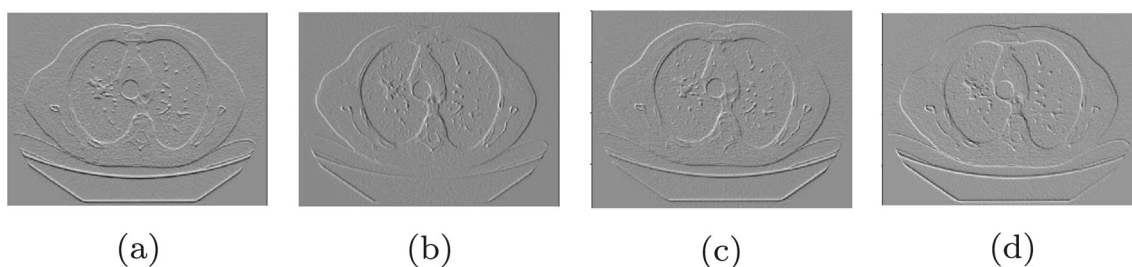


Fig. 3 **a** Low-dose CT image. Output of edge detection layer in **b** horizontal direction, **c** vertical direction, **d** 45° diagonal direction, and **e** 135° diagonal direction

image are given to a pre-trained convolutional network, one at a time. Then, the comparison is made between the feature maps generated by the two images. VGG-16 [33] is a pre-trained network for classification on ImageNet dataset [45] which is generally used to calculate the perceptual loss in generative adversarial networks.

In this study, we have incorporated both MSE and perceptual loss to optimize the network. Our experiments showed that using the perceptual loss solely results in a grid-like artifact in the output image. This effect has been perceived by other researchers, too [44]. Therefore, we have combined both per-pixel loss \mathcal{L}_{mse} and perceptual loss \mathcal{L}_P to advance the optimization.

$$\mathcal{L}(\theta) = \lambda_{mse}\mathcal{L}_{mse}(\theta) + \lambda_P\mathcal{L}_P(\theta) \tag{7}$$

where, λ_{mse} and λ_P are weighting scalars for mean-square error loss and perceptual loss, respectively. The optimal weights are found using validation dataset in each experiment. The mean-square error between the ground-truth y and the denoised image from the proposed network \hat{y} is defined as

$$\mathcal{L}_{mse}(\theta) = ||\hat{y}(\theta) - y||^2$$

Similar to other studies, we have employed VGG-16 network to measure the perceptual loss. The output of our proposed network is given to the VGG16 network and four groups of feature maps are extracted from it in different layers. These feature maps are compared with the feature maps that normal-dose CT image generates and then Eq. 8 is used to compute the perceptual loss. As Fig. 4 demonstrates, we have used the output of last convolutional layer (after ReLU activation and before pooling layer) in blocks 1, 2, 3, and 4. The perceptual loss function $\mathcal{L}_P(\theta)$ is

$$\mathcal{L}_P(\theta) = \sum_{i=1}^4 \mathcal{L}_i(\theta) \tag{8}$$

$$\mathcal{L}_i(\theta) = \frac{1}{h_i w_i d_i} ||\phi_i(\hat{y}(\theta)) - \phi_i(y)||^2 \tag{9}$$

here, ϕ_i refers to the extracted feature maps from block i with size $h_i \times w_i \times d_i$. Since VGG16 is a fully convolutional neural network, we are not limited to a specific input image

size, so h_i and w_i depend on the size of the input image. The number of filters in the convolutional layer defines the parameter d_i and is stated in [33].

Our experiments reveal that utilizing perceptual loss with the mean-square error greatly improves the visual characteristics of the output image.

Experiments Setup

In this study, we have used three datasets to evaluate the performance of the proposed network in removing noise from low-dose CT images: simulated dataset, real piglet dataset, and Thoracic dataset.

To create the simulated dataset, we downloaded lung CT scans [46] for a patient including 663 slices from The Cancer Imaging Archive (TCIA) [47]. The CT images were taken with 330 *mAs* X-ray current tube, 120 *KVp* peak voltage and 1.25 *mm* slice thickness. Then, with the procedure explained in “Low-Dose CT Simulation” generated low-dose CT images. The incident flux of simulated low-dose CT (I_{id}^0) in Eq. 3 is define equal to 2×10^3 .

The second dataset is a real dataset acquired from a deceased piglet. The dataset is produced by the authors of [29] and is available at [48] for download. It contains 900 images with 100 *KVp*, 0.625 *mm* thickness. The X-ray currents for normal-dose and low-dose images are 300 *mAs* and 15 *mAs*, respectively.

Thoracic dataset [49] includes 407 pairs of CT image from an anthropomorphic thoracic phantom. The current tube for normal-dose and low-dose CT images are 480 *mAs* and 60 *mAs*, respectively with a peak voltage of 120 *KVp* and slice thickness of 0.75 *mm*.

In each dataset, 60% of the images are used for training the proposed network, 20% for validation, and 20% for testing. Opposite to other studies that build a test dataset randomly, our test dataset holds the last 20% of images in the original dataset. The reason is that the consecutive CT images are very similar to each other and testing the network on the random dataset does not clearly examine the effectiveness of the network on the new images. Using

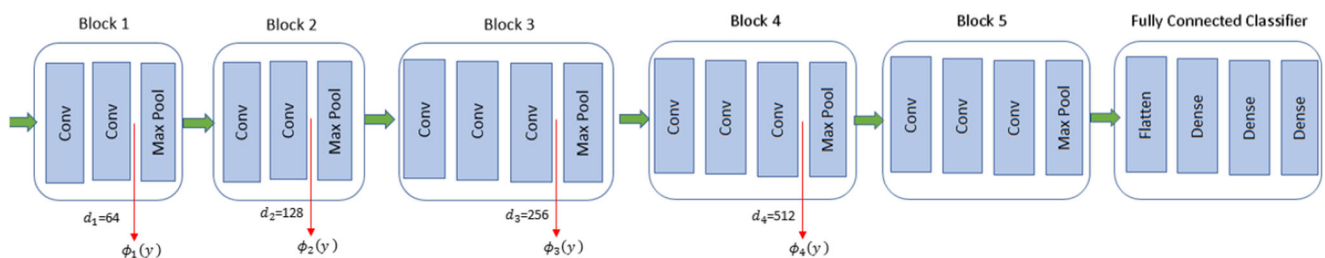


Fig. 4 Perceptual loss is computed by extracting the feature maps of blocks 1, 2, 3, and 4 from a pre-trained VGG-16 network

Table 2 The average PSNR and SSIM of the different algorithms for the Lung dataset

| Metric | Low-dose image | BM3D | CNN200 [20] | [41] | DRL | DRL-E-M | DRL-E-P | DRL-E-MP |
|--------|----------------|--------|-------------|--------|--------|---------|---------|----------|
| PSNR | 14.59 | 24.76 | 33.19 | 33.74 | 34.17 | 36.64 | 33.47 | 35.57 |
| SSIM | 0.2729 | 0.6750 | 0.6654 | 0.7739 | 0.8364 | 0.9033 | 0.7732 | 0.8463 |

the last portion of CT images assures us that the testing is performed on images that the network has not seen before. To prepare the data for training the network, we have used pixel values of low-dose, normal-dose images divided by 4095. This maps the data between 0 and 1 which is suitable for training neural networks.

The original size of CT images in all the mentioned datasets is 512×512 . To boost the number of training samples, we have extracted overlapping patches of 40×40 from images, as the receptive field of the proposed network is $5 + 4 + 6 + 8 + 6 + 4 + 2 + 2 = 37$ in each direction. This also helps to reduce the memory resources needed during training. Since the network is fully convolutional, the input size does not have to be fixed; test images with their original size are fed to the network. To avoid boundary artifacts, zero padding in convolutional operators is used [41]. The

activation function is rectified linear unit, and the number of filters in all convolutional layers is 64 except layers 7 and 8 which have 1 filter. To see how adding the edge detection layer and utilizing MSE and perceptual loss improve the performance, we have trained multiple networks. Training of all the networks is performed with Adam optimizer in two stages: with learning rate $1e - 3$ for 20 epochs and then learning rate $1e - 4$ for 20 more epochs which let them converge to the optimal solution. Glorot normal is used to initialize the weights [50]. Two more networks proposed in studies [20] and [41] have also been trained for comparison. The implementation was based on Keras with TensorFlow backend on the system with an Intel Core i7 CPU 3.4 GHz, 32 G memory and GeForce GTX 1070 Graphics Card. The code is available at <https://github.com/maransari/low-dose-CT-denoising>.

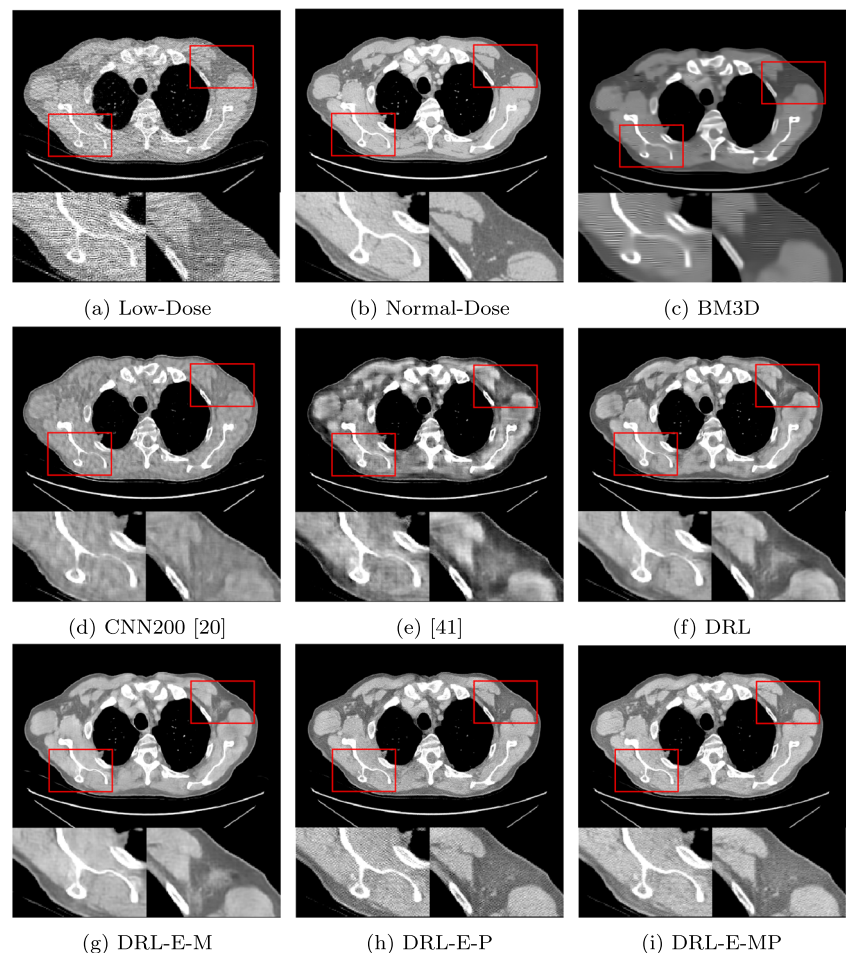
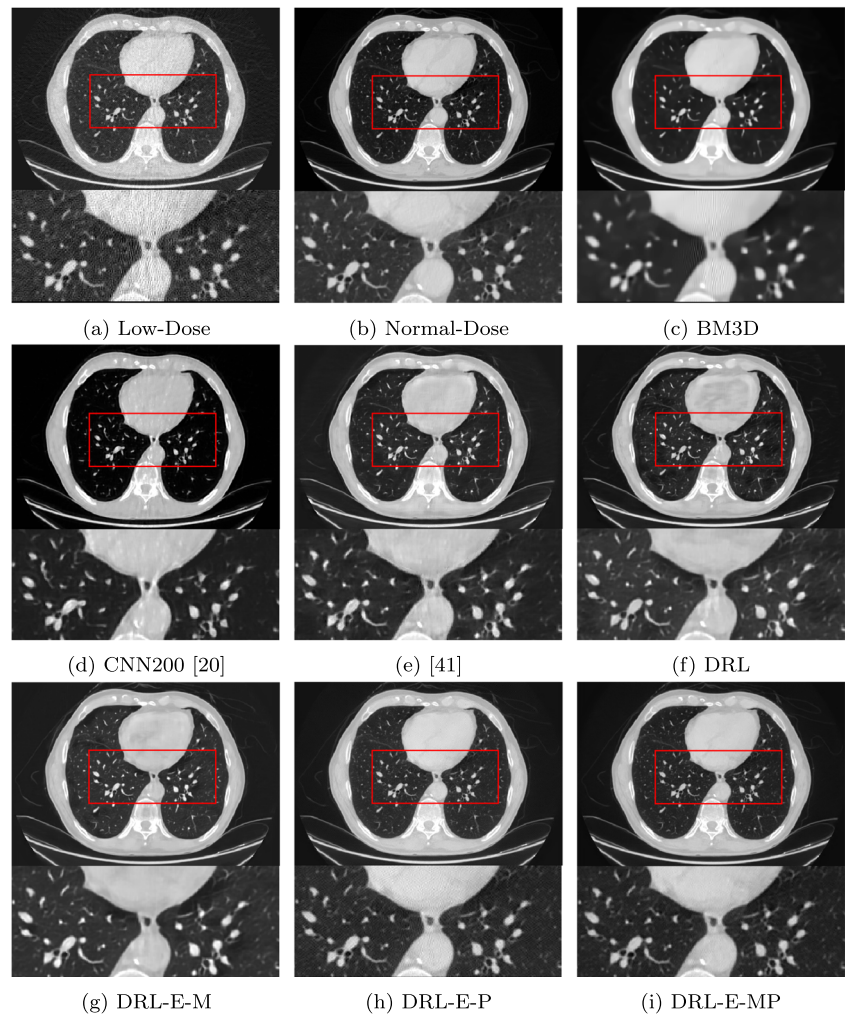
Fig. 5 Denoising results of different algorithms on Lung dataset in abdomen window

Fig. 6 Denoising results of the different algorithms on Lung dataset in lung window



Results

To evaluate the performance of the proposed network, we have compared the results with state-of-the-art BM3D [15], and two neural networks, CNN200 [20] and [41]. These networks are retrained over each dataset. As mentioned earlier, the initial idea of the proposed network was derived from [41] designed for image super-resolution. To investigate how each change in the network architecture affects the performance, we have made three more comparisons with three more networks.

The first network is designed to examine how residual learning enhances the outcome. This network is similar to the one in [41], but there are shortcuts between the outputs

of layers 2 and 3 with the outputs of layers 6 and 5, respectively. We call this network DRL (dilated residual learning). The objective function for this network is the mean-square error (MSE), and we demonstrate that adding shortcut connections improves the results.

In the second network, we have added the edge detection layer to the beginning of the network. This network is named DRL-E and is shown in Fig. 1. We have optimized this network by three objective functions to investigate the effects of choosing a loss function on the results. First, this network is optimized by MSE loss function and we refer to it as DRL-E-M. Next, we optimized the network by perceptual loss and it is called DRL-E-P in this paper. Finally, the proposed network is trained by the objective function

Table 3 The average PSNR and SSIM of the different algorithms for the Piglet dataset

| Metric | Low-dose image | BM3D | CNN200 [20] | [41] | DRL | DRL-E-M | DRL-E-P | DRL-E-MP |
|--------|----------------|--------|-------------|--------|--------|---------|---------|----------|
| PSNR | 39.93 | 41.46 | 44.18 | 44.83 | 44.96 | 45.10 | 44.01 | 44.12 |
| SSIM | 0.9168 | 0.9733 | 0.9804 | 0.9816 | 0.9881 | 0.9885 | 0.9782 | 0.9807 |

defined in Eq. 7. This network optimizes a combination of mean-square error and perceptual loss to learn the weights and achieve the best results. We refer to this combination as DRL-E-MP.

For each algorithm, we have provided the quantitative results proving that the proposed network DRL-E with dilated convolutions, shortcut connections, and the edge detection layer outperforms the other networks. Moreover, visual comparisons confirm that utilizing the proposed objective function improves the perceptual aspects of the DRL-E network further and conserves most of the details in the image.

Denoising Results on Simulated Lung Dataset

Table 2 displays the average peak signal to noise ratio (PSNR) and structural similarity (SSIM) of applying the state-of-the-art BM3D and six neural networks. Figures 5 and 6 give the visual results for the Lung dataset in two different windows. Windowing helps to visualize the details of CT images properly. Here, we have shown the results in the lung and abdominal window to distinguish the differences better. Abdomen window helps to distinguish small changes in density and displays more texture details. Since the lung is air-filled, it has very low density and

Fig. 7 Denoising results of the different algorithms on Piglet dataset in abdomen window

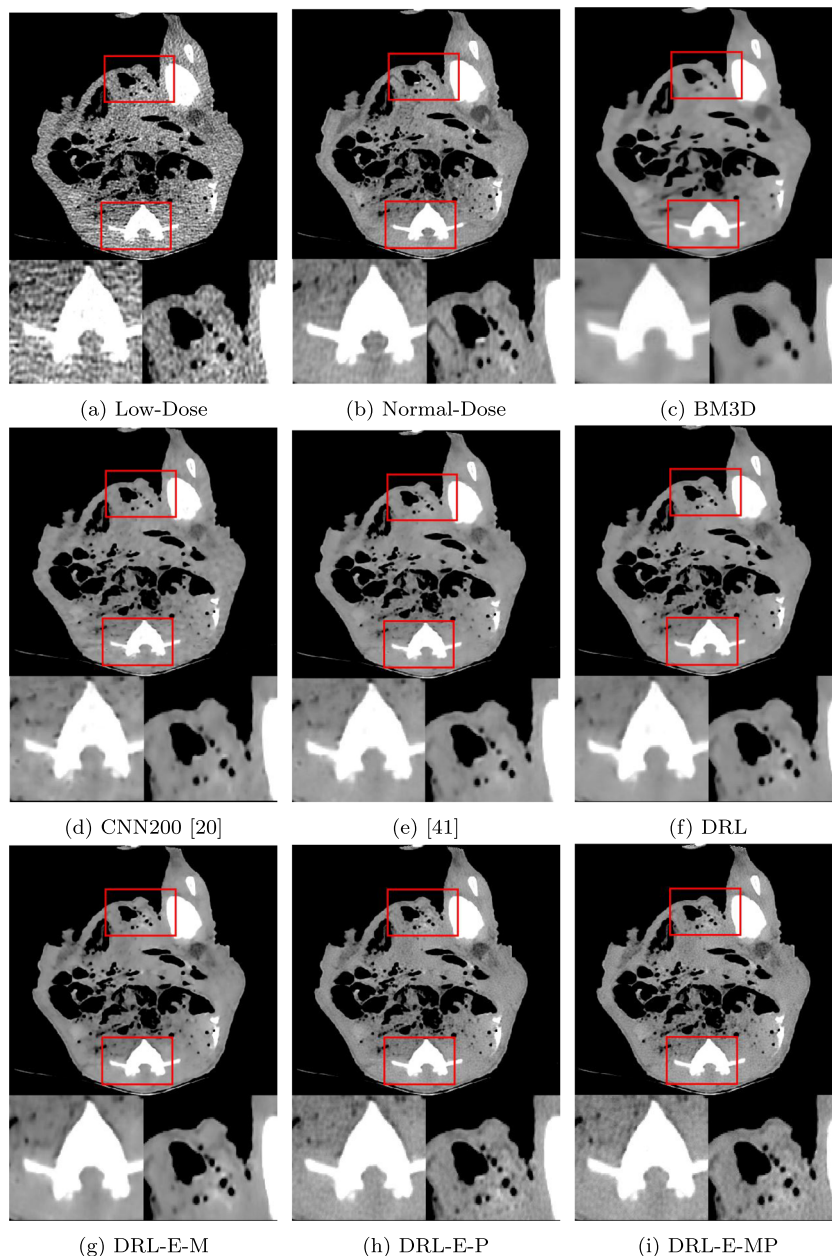


Table 4 The average PSNR and SSIM of the different algorithms for the Thoracic dataset

| Metric | Low-dose image | BM3D | CNN200 [20] | [41] | DRL | DRL-E-M | DRL-E-P | DRL-E-MP |
|--------|----------------|--------|-------------|--------|--------|---------|---------|----------|
| PSNR | 25.66 | 30.86 | 33.57 | 33.73 | 34.02 | 34.03 | 26.25 | 31.50 |
| SSIM | 0.3711 | 0.6252 | 0.6614 | 0.6761 | 0.6674 | 0.6697 | 0.3673 | 0.5485 |

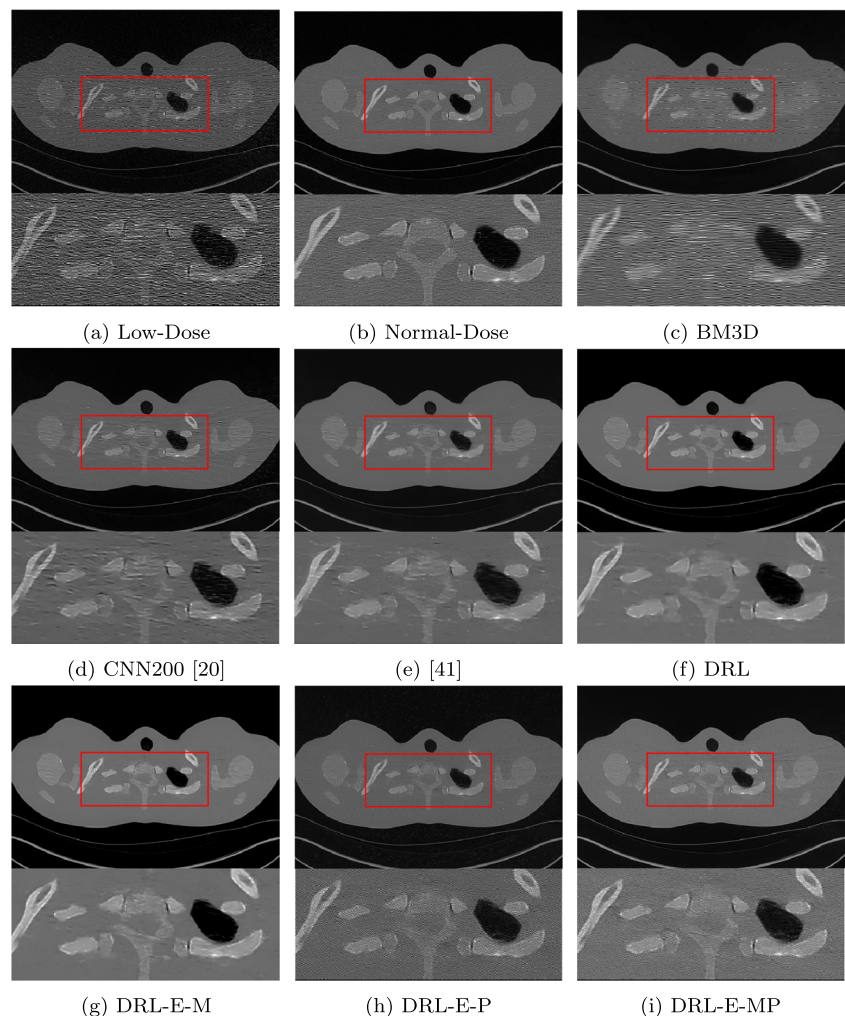
appears black in the abdomen window. Lung window improves the visibility of the areas of consolidation and pulmonary vascular structures.

Figures 5 and 6 approve that the alterations of the network have enhanced the outcome step by step. By comparing Fig. 5f and g, we can see that adding the edge detection layer generates sharper and more distinct edges. This improvement also can be concluded from Table 2. When the objective function is mean-square error, the network with edge detection layer offers higher PSNR and SSIM. This dataset also demonstrates the effectiveness of perceptual loss. The results show using MSE as an objective function generates smooth regions and effects the details in the texture. On the other hand, perceptual loss forces

the output of the network to be perceptually similar to the ground truth. However, training the network solely by perceptual loss generates grid-like artifacts in the output image. As the results of DRL-E-MP demonstrates, the combined objective function saves most of the details in the textures and provides a better visual outcome.

As one can expect, exploiting perceptual loss do not improve PSNR. The reason is that high PSNR is the result of minimizing per-pixel loss. If a network is trained to minimize MSE, it will always have higher PSNR compare to a network that is trained to minimize the perceptual loss. However, as we can see here, higher PSNR does not always provide the most visually appealing outcome. Similar results have been observed in [44].

Fig. 8 Denoising results of the different algorithms on Thoracic dataset in abdomen window



Denoising Results on Real Piglet Dataset

Table 3 displays the quantitative effects of performing the denoising on real low-dose CT images for Piglet dataset which approves the results obtained from simulated Lung dataset. Comparing the PSNR of the BM3D, CNN200, [41], DRL, and DRL-E demonstrates that when the objective function is MSE, the network with residual learning and edge detection layer outperforms the other ones. Figure 7 provides a visual comparison between the outcomes. It reveals that joining perceptual loss and per-pixel loss further improves the produced images by the proposed network. The smoothing effect of MSE optimization helps to discard the grid-like artifacts caused by perceptual loss. DRL-E-MP resembles the normal-dose CT image better by reconstructing fine details.

Denoising results on phantom Thoracic dataset

Table 4 represents the PSNR and SSIM of denoising Thoracic dataset by all the methods. Results obtained for this dataset is consistent with the other experiments. Figure 8 clearly exhibits the effects of each alteration. Comparing the results obtained by DRL and DRL-E-M confirms that the edge detection layer helps to deliver sharper and more precise edges. As explained before, the only difference between these two models is using the edge detection layer.

Conclusion

In this paper, we have combined the benefits of dilated convolution, residual learning, edge detection layer, and perceptual loss to design a noise removal deep network that produces normal-dose CT image from low-dose CT image. First, we have designed a network by the adoption of dilated convolution instead of standard convolution and also, using residual learning by adding symmetric shortcut connections. We have, also, implemented an edge detection layer that acts the same as Sobel operator and helps to capture the boundaries in the image better. In the case of the objective function, we have observed that optimizing by a joint function of MSE loss and perceptual loss provides better visual results compared to each one alone. The obtained results do not suffer from over smoothing and loss of details that are the results of per-pixel optimization and the grid-like artifacts occurring with perceptual loss optimization.

Acknowledgements This work was supported in part by a research grant from Natural Sciences and Engineering Research Council of Canada (NSERC). The authors would like to thank Dr. Paul Babyn and Troy Anderson for the acquisition of the piglet dataset. The results

shown here are in whole or part based upon data generated by the TCGA Research Network: <http://cancergenome.nih.gov/>.

Funding Information This work was supported in part by a research grant from Natural Sciences and Engineering Research Council of Canada (NSERC).

References

- Bencardino J T: Radiological society of north america (rsna) 2010 annual meeting. *Skelet Radiol* 40:1109–1112, 2011
- Donya M, Radford M, ElGuindy A, Firmin D, Yacoub M H (2015) Radiation in medicine: origins, risks and aspirations. *Global Cardiology Science and Practice* pp 57
- Ehman E C, Yu L, Manduca A, Hara A K, Shiung M M, Jondal D, Lake D S, Paden R G, Blezek D J, Bruesewitz M R, et al: Methods for clinical evaluation of noise reduction techniques in abdominopelvic CT. *Radiographics* 34(4):849–862, 2014
- Wang J, Lu H, Liang Z, Eremina D, Zhang G, Wang S, Chen J, Manzione J: An experimental study on the noise properties of x-ray CT sinogram data in radon space. *Phys Med Biol* 53(12):3327, 2008
- Macovski A: *Medical Imaging Systems*, vol 20 NJ: Prentice-Hall Englewood Cliffs, 1983
- Manduca A, Yu L, Trzasko J D, Khaylova N, Kofler J M, McCollough C M, Fletcher J G: Projection space denoising with bilateral filtering and CT noise modeling for dose reduction in CT. *Med Phys* 36(11):4911–4919, 2009
- Wang J, Li T, Lu H, Liang Z: Penalized weighted least-squares approach to sinogram noise reduction and image reconstruction for low-dose x-ray computed tomography. *IEEE Trans Med Imaging* 25(10):1272–1283, 2006
- Pickhardt P J, Lubner M G, Kim D H, Tang J, Ruma J A, del Rio A M, Chen G H: Abdominal CT with model-based iterative reconstruction (mbir): initial results of a prospective trial comparing ultralow-dose with standard-dose imaging. *Am J Roentgenol* 199(6):1266–1274, 2012
- Fletcher J G, Grant K L, Fidler J L, Shiung M, Yu L, Wang J, Schmidt B, Allmendinger T, McCollough C H: Validation of dual-source single-tube reconstruction as a method to obtain half-dose images to evaluate radiation dose and noise reduction: phantom and human assessment using CT colonography and sinogram-affirmed iterative reconstruction (safire). *J Comput Assist Tomogr* 36(5):560–569, 2012
- Aharon M, Elad M, Bruckstein A et al: K-svd: an algorithm for designing overcomplete dictionaries for sparse representation. *IEEE Trans Signal Process* 54(11):4311, 2006
- Chen Y, Yin X, Shi L, Shu H, Luo L, Coatrieux J L, Toumoulin C: Improving abdomen tumor low-dose CT images using a fast dictionary learning based processing. *Phys Med Biol* 58(16):5803, 2013
- Abhari K, Marsousi M, Alirezaie J, Babyn P (2012) Computed tomography image denoising utilizing an efficient sparse coding algorithm. 2012 11th International Conference on Information Science, Signal Processing and their Applications (ISSPA) pp 259–263
- Buades A, Coll B, Morel J M (2005) A non-local algorithm for image denoising. In: *IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2005. CVPR 2005*, vol. 2, pp 60–65. IEEE
- Chen Y, Yang Z, Hu Y, Yang G, Zhu Y, Li Y, Chen W, Toumoulin C et al: Thoracic low-dose CT image processing using

- an artifact suppressed large-scale nonlocal means. *Phys Med Biol* 57(9):2667, 2012
15. Dabov K, Foi A, Katkovnik V, Egiazarian K: Image denoising by sparse 3-d transform-domain collaborative filtering. *IEEE Trans Signal Process* 16(8):2080–2095, 2007
 16. Hashemi S, Paul N S, Beheshti S, Cobbold R S (2015) Adaptively tuned iterative low dose CT image denoising. *Computational and mathematical methods in medicine* pp 2015
 17. Kang D, Slomka P, Nakazato R, Woo J, Berman D S, Kuo C C J, Dey D: Image denoising of low-radiation dose coronary CT angiography by an adaptive block-matching 3d algorithm. In: *Medical imaging 2013: Image processing*, vol. 8669, p. 86692g. International society for optics and photonics, 2013
 18. Ioffe S, Szegedy C: Batch normalization: accelerating deep network training by reducing internal covariate shift. In: *ICML*, 2015
 19. He K, Zhang X, Ren S, Sun J (2016) Deep residual learning for image recognition. *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* pp 770–778
 20. Chen H, Zhang Y, Zhang W, Liao P, Li K, Zhou J, Wang G: Low-dose CT via convolutional neural network. *Biomed Opt Express* 8(2):679–694, 2017
 21. Dong C, Loy C C, He K, Tang X: Image super-resolution using deep convolutional networks. *IEEE Trans Pattern Anal Mach Intell* 38(2):295–307, 2016
 22. Nishio M, Nagashima C, Hirabayashi S, Ohnishi A, Sasaki K, Sagawa T, Hamada M, Yamashita T: Convolutional auto-encoder for image denoising of ultra-low-dose CT. *Heliyon* 3(8):e00393, 2017
 23. Chen H, Zhang Y, Kalra M K, Lin F, Chen Y, Liao P, Zhou J, Wang G: Low-dose CT with a residual encoder-decoder convolutional neural network. *IEEE Trans Med Imaging* 36(12):2524–2535, 2017
 24. Kang E, Min J, Ye J C (2017) A deep convolutional neural network using directional wavelets for low-dose x-ray CT reconstruction. *Medical physics* 44(10)
 25. Goodfellow I J, Pouget-Abadie J, Mirza M, Xu B, Warde-Farley D, Ozair S, Courville A C, Bengio Y (2014) Generative adversarial networks. arXiv:1406.2661
 26. Reed S, Akata Z, Yan X, Logeswaran L, Schiele B, Lee H (2016) Generative adversarial text to image synthesis. arXiv:1605.05396
 27. Ledig C, Theis L, Huszár F, Caballero J, Cunningham A, Acosta A, Aitken A P, Tejani A, Totz J, Wang Z, et al: Photo-realistic single image super-resolution using a generative adversarial network. In: *CVPR*, vol 2, p 4, 2017
 28. Vondrick C, Pirsivash H, Torralba A: Generating videos with scene dynamics. In: *Advances in neural information processing systems*, pp 613–621, 2016
 29. Yi X, Babyn P (2018) Sharpness-aware low-dose CT denoising using conditional generative adversarial network. *Journal of digital imaging*, pp 1–15
 30. Wolterink J M, Leiner T, Viergever M A, Išgum I: Generative adversarial networks for noise reduction in low-dose CT. *IEEE Trans Med Imaging* 36(12):2536–2545, 2017
 31. Yang Q, Yan P, Zhang Y, Yu H, Shi Y, Mou X, Kalra M K, Zhang Y, Sun L, Wang G (2018) Low dose CT image denoising using a generative adversarial network with wasserstein distance and perceptual loss. *IEEE transactions on medical imaging*
 32. Yang Q, Yan P, Kalra M K, Wang G (2017) CT image denoising with perceptive deep neural networks. arXiv:1702.07019
 33. Simonyan K, Zisserman A (2014) Very deep convolutional networks for large-scale image recognition. arXiv:1409.1556
 34. Bevens N, Szczykutowicz T, Supanich M: Tu-c-103-06: a simple method for simulating reduced-dose images for evaluation of clinical CT protocols. *Med Phys* 40(6Part26):437–437, 2013
 35. Zeng D, Huang J, Bian Z, Niu S, Zhang H, Feng Q, Liang Z, Ma J: A simple low-dose x-ray CT simulation from high-dose scan. *IEEE Trans Nucl Sci* 62(5):2226–2233, 2015
 36. Yu F, Koltun V (2015) Multi-scale context aggregation by dilated convolutions. arXiv:1511.07122
 37. Chen L C, Papandreou G, Kokkinos I, Murphy K, Yuille A L: Deeplab: semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected crfs. *IEEE Trans Pattern Anal Mach Intell* 40(4):834–848, 2018
 38. Mao X, Shen C, Yang Y B: Image restoration using very deep convolutional encoder-decoder networks with symmetric skip connections. In: *Advances in neural information processing systems*, pp 2802–2810, 2016
 39. Wang T, Sun M, Hu K: Dilated deep residual network for image denoising. In: *2017 IEEE 29th international conference on tools with artificial intelligence (ICTAI)*, pp 1272–1279. IEEE, 2017
 40. Zhang K, Zuo W, Chen Y, Meng D, Zhang L (2017) Beyond a gaussian denoiser: residual learning of deep cnn for image denoising. *IEEE Transactions on Image Processing*
 41. Zhang K, Zuo W, Gu S, Zhang L: Learning deep cnn denoiser prior for image restoration. In: *IEEE Conference on Computer Vision and Pattern Recognition*, vol. 2, 2017
 42. Huang G, Liu Z, Weinberger KQ, van der Maaten L (2016) Densely connected convolutional networks. arXiv:1608.06993
 43. Sobel I (1990) An isotropic 3 × 3 image gradient operator. *Machine vision for three-dimensional scenes* pp 376–379
 44. Johnson J, Alahi A, Fei-Fei L: Perceptual losses for real-time style transfer and super-resolution. In: *European Conference on Computer Vision*, pp 694–711. Springer, 2016
 45. Deng J, Dong W, Socher R, Li L J, Li K, Fei-Fei L: Imagenet: a large-scale hierarchical image database. In: *CVPR09*, 2009
 46. Lingle W, Erickson B, Zuley M, Jarosz R, Bonaccio E, Filippini J, Grusauskas N (2016) Radiology data from the cancer genome atlas breast invasive carcinoma [tcga-brca] collection. *The Cancer Imaging Archive*
 47. Clark K, Vendt B, Smith K, Freymann J, Kirby J, Koppel P, Moore S, Phillips S, Maffitt D, Pringle M, et al: The cancer imaging archive (tcia): maintaining and operating a public information repository. *J Digit Imaging* 26(6):1045–1057, 2013
 48. Yi X (2019) Recent publication. <http://homepage.usask.ca/xiy525/>
 49. Gavrielides MA, Kinnard LM, Myers KJ, Peregoy J, Pritchard WF, Zeng R, Esparza J, Karanian J, Petrick N: A resource for the assessment of lung nodule size estimation methods: database of thoracic ct scans of an anthropomorphic phantom. *Opt Express* 18(14):15,244–15,255, 2010. <https://doi.org/10.1364/OE.18.015244>. <http://www.opticsexpress.org/abstract.cfm?URI=oe-18-14-15244>
 50. Glorot X, Bengio Y: Understanding the difficulty of training deep feedforward neural networks. In: *Proceedings of the 13th International Conference on Artificial Intelligence and Statistics*, pp 249–256, 2010