

Low Dose CT Image Denoising using Deep Convolutional Neural Networks with Extended Receptive Fields

This paper was downloaded from TechRxiv (<https://www.techrxiv.org>).

LICENSE

CC BY 4.0

SUBMISSION DATE / POSTED DATE

31-10-2021 / 04-11-2021

CITATION

Thanh-Trung, Nguyen; Trinh, Dinh-Hoan; Linh-Trung, Nguyen; Luong, Marie (2021): Low Dose CT Image Denoising using Deep Convolutional Neural Networks with Extended Receptive Fields. TechRxiv. Preprint. <https://doi.org/10.36227/techrxiv.16910395.v1>

DOI

[10.36227/techrxiv.16910395.v1](https://doi.org/10.36227/techrxiv.16910395.v1)

Low-Dose CT Image Denoising using Deep Convolutional Neural Networks with Extended Receptive Fields

Nguyen Thanh-Trung^{1,3}, Dinh-Hoan Trinh^{1,2}, *Nguyen Linh Trung¹, Marie Luong⁴

¹Advanced Institute of Engineering and Technology, University of Engineering and Technology, Vietnam National University, Hanoi, Vietnam

²Viettel Cyberspace Center, Viettel Group, Vietnam

³University of Information and Communication Technology, Thai Nguyen University, Thai Nguyen, Vietnam

⁴L2TI Laboratory, Galillee Institute, University Paris 13, France

Correspondence to: Dinh-Hoan Trinh (hoantd5@viettel.com.vn or tdhoan@vnu.edu.vn); Nguyen Linh Trung (linhtrung@vnu.edu.vn)

Received: 7 October 2021 / Accepted: date

Abstract How to reduce dose radiation while preserving the image quality as when using standard dose is one of the most important topics in the Computed Tomography (CT) imaging domain due to quality of low dose CT (LDCT) images is often strongly affected by noise and artifacts. Recently there has been considerable interest in using deep learning as a post-processing step to improve the quality of reconstructed LDCT images. This paper, first, gives an overview of learning-based LDCT image denoising methods from patch-based early learning methods to state-of-the-art CNN-based ones, and then a novel CNN-based method is presented. In the proposed method, preprocessing and post-processing techniques are integrated into a dilated convolutional neural network to extend receptive fields. Hence, large distance pixels in input images will participate in enriching feature maps of the learned model and thus effectively denoises. Experimental results showed that the proposed method is light while its denoising effectiveness is competitive with well-known CNN-based models.

1 Introduction

The presence of noise adversely affects the image quality as well as the performance of subsequent image analysis and processing tasks. Thus, denoising always plays an important role in modern image processing systems. Although image denoising has been studied for a long time with numerous efficient methods, it still remains an active area of research as it is the testbed for a variety of high level image processing tasks. In this paper, we are interested in one of the most difficult denoising problems, namely denoising for low-dose Computed Tomography (LDCT) images. X-Ray CT, introduced in the early 1970s [1], is an ionizing radiation-based medical imaging technique. CT has been one of the most widely used imaging modalities in medicine. However, its major drawback is that the use of ionizing radiation can be harmful to the health of the patient. Therefore, it is necessary to reduce radiation dose. Unfortunately, it was showed

that the reduction of radiation dose leads to increased noise level and artifacts in reconstructed images. These factors affect the quality of CT images and thereby the diagnostic accuracy and the outcome of a CT examination. Thus, the reconstruction of high-quality CT images under low/ultra-low radiation dose conditions is a great challenge.

Numerous studies have been carried out for denoising and artifact removal for low-dose CT. The existing methods can be classified into two groups: (i) the group of methods which filter noise and reduce artifacts within the image reconstruction from raw projection data, and (ii) the group of methods which perform removing noise and artifacts in reconstructed low-dose CT images. For the first group, methods mostly use filters to suppress noise and artifacts in the raw data (sinogram). The filtered sinogram data are then used to reconstruct the image using an iterative (IR) or filtered back-projection (FBP) method [2, 4, 5]. In the filters, the noise prior (e.g., Poisson noise) plays an important role in designing denoising algorithms. In fact, exactly estimating the noise distribution and level is not easy. It is shown that sinogram filtration and IR techniques can reconstruct low-dose CT images with quality equivalent to current clinical standards [6]. It is noticed that the reconstruction techniques are embedded in the hardware of CT scan systems and thus the sinogram data are not available to users. For the second group (group of post-processing methods which do not rely on raw data), numerous digital image denoising methods were modified to be able to apply to low-dose CT images such as the nonlocal means filter-based methods [7–9], the methods using sparse representation [10–12], the filters in the Wavelet domain [13, 14], the BM3D filter [15]. For these classical methods, noise is directly suppressed from the noisy image. Similar to sinogram filtration, they require the knowledge about the types of source noise and general properties of noise in CT images. Most existing methods assume that the distribution of the noise in CT images can be approximated by the Gaussian distribution [16].

Trinh *et al.* in [17] proposed to use the assumption of local Gaussian distribution in dealing with the variation of noise levels among image regions. However, the fact that the noise in low-dose CT is very complex makes it difficult to be exactly estimated, affecting the LDCT image denoising performance of the classical methods.

Recently, deep Convolutional Neural Networks (CNN) have been used to denoise LDCT images with impressive results. It is shown that the CNN-based methods, for instances, RED-CNN [18], WGAN-VGG [19], SAGAN [20], SMGAN [21], FD-VGG [22], significantly outperform classical denoising methods. In this post-processing approach, a large dataset of low-dose and normal-dose image pairs (images of such a pair are taken at the same position of a patient) is required. Normal-dose CT (NDCT) images are much less noisy and have higher quality, as compared to LDCT images. The aim of CNN-based methods is to train a deep network to learn the mapping from the LDCT image to the NDCT image of a pair. However, the idea of using high quality CT images for denoising LDCT images was early proposed [23–25]. An overview of learning-based LDCT image denoising methods, thus, is necessary.

In this paper, we first overview existing learning-based methods for LDCT image denoising from naive methods to state-of-the-art ones. Typical learning-based methods will be subjectively and objectively compared through experiments performed on LDCT images. Then, we will present a novel CNN-based method that uses an extended receptive field CNN architecture. In this method, noisy images are re-arranged in an orderly manner into sequences of subimages (pre-processing). A deep dilated residual CNN network receives these subimage sequences and learns the mapping from LDCT to NDCT images in a given training dataset. At the end of the network, a reconstruction stage is used to reconstruct the desired output image (post-processing). By using the dilated convolution and the pre-processing step, the receptive fields in the network are extended so as to be able to take into account more useful information for denoising.

The rest of the paper is organized as follows. In Section 2, an overview of existing example-based learning methods for LDCT image denoising will be given, with details about the main contents of the typical methods. The proposed novel contribution is presented in Section 3. Objective and subjective comparisons are reported in Section 4. Finally, conclusion is given in the last section.

2 Overview of Learning-based Methods for LDCT Image Denoising

Generally, the goal of image denoising is to restore a clean image from its noisy observations. Unlike the classical methods, which try to directly denoise by solving a inverse problem models (e.g., sparse representation, statistical filters, total variation), learning-based methods learn the mapping that represents the relationship between noisy and clean image

pairs from given external datasets and use the trained mapping to denoise the new noisy images. In this section, the main contributions in this approach of denoising and their applications to LDCT images (from the naive methods to the current state-of-the-art methods) will be mentioned. Before going into the details of the existing learning-based methods, let us start with the external patch-based denoising methods which are considered as a bridge between classical methods and learning-based methods.

2.1 External Patch-based Denoising Methods

Recent CNN-based methods [18, 19, 22] are known as the state-of-the-arts for LDCT image denoising. However, the idea of using NDCT images to denoise LDCT images was early proposed in the patch-based denoising methods [17, 23, 25, 26]. Although computation time is a drawback of these patch-based methods, their denoising effectiveness has been confirmed. This approach of external patch-based denoising can be considered as a bridge from classical denoising methods to learning-based methods. So, this subsection briefly reviews some important contents of this approach.

The idea of an external patch-based denoising algorithm can be formulated as follows. A large noisy image is considered as an arranged set of overlapped small patches, and denoising is performed on patches. Given a patch of size $\sqrt{n} \times \sqrt{n}$ from a noisy image, presented by a vector $\mathbf{q} \in \mathcal{R}^n$, the algorithm finds a set of similar (i.e., reference) patches $\mathbf{p}_1, \mathbf{p}_2, \dots, \mathbf{p}_k \in \mathcal{R}^n$ from external clean images and determines a mapping \mathcal{F} to obtain an estimate $\hat{\mathbf{p}}$ of the unknown clean patch \mathbf{p} as

$$\hat{\mathbf{p}} = \mathcal{F}(\mathbf{p}_1, \mathbf{p}_2, \dots, \mathbf{p}_k). \quad (1)$$

Numerous patch-based image denoising methods using external clean images have been proposed [17, 25, 26]. In an NLM algorithm proposed in [26], the denoising function \mathcal{F} of (1) is defined based on the weighted average model of the nonlocal mean method [27]. However, the weights are computed using reference patches extracted from NDCT images. Trinh *et al.* in [17, 25] proposed to define \mathcal{F} as a sparse linear combination of reference patches, to be determined by solving non negative sparse coding problem. These methods demonstrated that, by using external patches, noise in LDCT images can be effectively suppressed while preserving subtle details, as compared to classical methods. Nguyen *et al.* in [28] used image decomposition and sparse representation techniques to define \mathcal{F} in such a way that it maximizes the preservation of high frequency components in LDCT images. Moreover, denoising with the help of external datasets of clean image patches for other image types has already studied in such as [29–31]. As an example, Lou *et al.* in [31] showed that, by using targeted external databases, one can obtain an effective denoising method, significantly outperformed well-known classical methods such as NLM [27], KSVD [10], and BM3D [15].

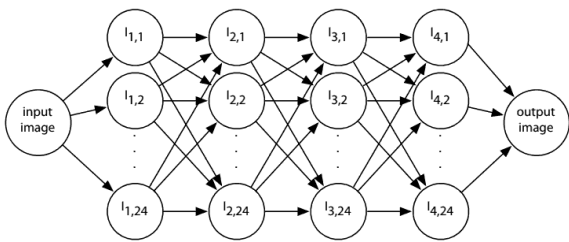


Fig. 1 One of the first CNN architectures for image denoising [32].

The advantage of this approach that it only needs to use clean patches in example datasets, it do not require the correspondences noise-free and noisy patch pairs as in the learning-based methods. Even though external patch-based denoising methods showed their effectiveness, this approach still requires an assumption of prior knowledge about the noise distribution. Global or local Gaussian noise is often used in most existing methods while noise in LDCT images is, in fact, very complex. Moreover, from model (1) one can see that the performance of external patch-based methods highly depends on the quality of found reference patches (candidate patches of the desired patch) and the mathematical model used to determine the denoising function \mathcal{F} . It is noticed that for every noisy patch the determination of reference patches and function \mathcal{F} is time-consuming. Thus, computational time is also a drawback of this approach.

2.2 Early Learning-based Denoising Methods

The goal of the learning-based denoising approach is to answer the question: Can we automatically learn a denoising procedure from given training examples that consist of pairs of noisy and noise-free and noiseless images? Specifically, a training dataset \mathbf{D} is defined as

$$\mathbf{D} = \{(\mathbf{n}_1, \mathbf{p}_1), (\mathbf{n}_2, \mathbf{p}_2), \dots, (\mathbf{n}_N, \mathbf{p}_N)\}, \quad (2)$$

where \mathbf{n}_i ($i = 1, 2, \dots, N$) are the noisy images, \mathbf{p}_i and their corresponding noise-free images. In LDCT image denoising, the training set is often established by the pair of low-dose (noisy) and normal-dose (noiseless) images, taken at the same position on the same patient. Learning-based methods try to train a model that represents the mapping from the space of noisy images to the space of clean images and use this trained model to remove noise from noisy images.

One of the first learning-based denoising methods was given in [32] by Jain and Seung. They designed simple convolutional neural networks of 4 layers, 24 neurons on each layer, and without activation functions (e.g., ReLU, Leaky ReLU) to build a denoising model for natural images (Fig. 1). Due to hardware limitation for training deep learning networks at the time and the simplicity of the network model, its denoising performance cannot be compared to the cleverly engineered algorithms (e.g., KSVD [10], BM3D [15]). Several other naive learning-based methods were proposed, for instances the nonlinear regression models in [23, 33].

Burger *et al.* in [34] proposed to use a deep learning model namely Multi-Layer Perceptron (MLP) and obtained the best performing method at that time. The authors also demonstrated that this learning-based approach can effectively denoise with and without known noise conditions.

Applications of example-based learning approaches to LDCT image denoising were early proposed in, for examples, [23] using kernel ridge regression, or in [35] using the Markov Random Field. Although simple learning models were used, the authors demonstrated promising results of the learning approach for LDCT image denoising.

In most of early learning-based methods, denoising a noisy image was performed on overlapping patches and final denoising results were aggregated from denoised patches. This way makes them difficult to control subtle textures at pixels in overlapped regions. However, the success of patch-based learning methods, e.g., the MLP method [34], opens a door for numerous CNN-based state-of-the-art denoising methods which are introduced in the next subsection.

2.3 CNN-based LDCT Denoising Methods

The outstanding success of numerous CNN-based denoising methods for natural images [36–39] and promising results of early learning-based methods for the problem of LDCT image denoising lead to many state-of-the-art CNN-based LDCT image denoising methods have been proposed in [18–21]. Most CNN-based methods were developed and experimented on open datasets such as the AAPM Low Dose CT Grand Challenge dataset [40] or, recently, the LoDoPaB-CT dataset [41]. In the following, we mention several recent typical CNN-based methods for LDCT image denoising.

RED-CNN [18], which originates from [42], is one of the first CNN-based methods for LDCT image denoising. The method uses the U-net architecture [43] in which pooling layers are replaced by convolutions and unpooling ones are replaced by deconvolutions. The convolutional layers (encoder) extract coarse features of the LDCT image, thus removing noise, while the deconvolutional layers (decoder) tend to recover subtle details which may be lost when the LDCT image is passed through the convolutional layers. The symmetric skip connections help the network converge faster and keep more subtle details [18]. RED-CNN can greatly reduce noise and artifacts. However, since RED-CNN only uses the mean-squared-error (MSE) loss function, the denoised images are often over-smoothed. The architecture of RED-CNN is shown in Fig. 2.

FD-VGG [22] is a solution that applies the FFD-net, designed in [37] for natural image denoising, to LDCT image denoising. The structure of the network is illustrated in Fig. 3. The authors proposed to use a loss function defined as the combination of MSE loss and “perceptual” loss in order to improve the global quality of denoised images. The perceptual loss was used in SACNN [44] with a self-attention

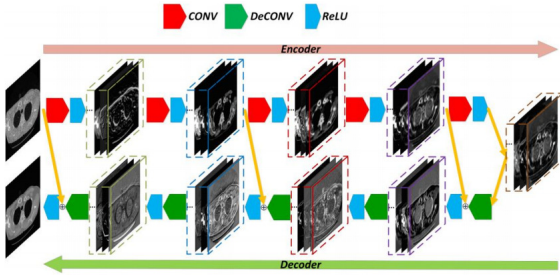


Fig. 2 Architecture of RED-CNN in [18].

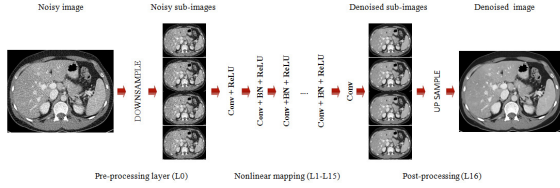


Fig. 3 Architecture of FD-VGG [22] inspired by the FFDNet method [37] for LDCT image denoising.

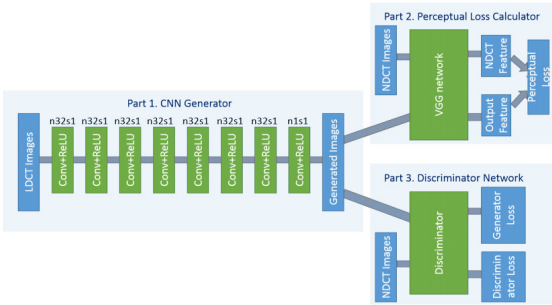


Fig. 4 GAN-based architecture of the WGAN-VGG method [19].

CNN. Shan *et al.* in [45] proposed a Convolutional Encoder-Decoder network with 2D and 3D configurations.

Other methods are based on Generative Adversarial Networks (GAN), being a deep-learning-based generative model introduced in 2014 by Goodfellow *et al.* [46]. The use of GAN for image denoising has been proposed in literature. In LDCT image denoising, state-of-the-art methods, e.g., WGAN [47], WGAN-VGG [19], SAGAN [20], SMGAN [21], demonstrated their high performance for estimating NDCT images from LDCT images. The architecture of GANs includes a generator and a discriminator. The generator is used to generate NDCT images from LDCT images and the discriminator is used to discriminate the generated (fake) NDCT images from real NDCT images. Du *et al.* in [48] applied GAN framework with a visual attention mechanism to better preserve details in denoised images.

As an example, Fig. 4 shows the architecture of WGAN-VGG [19], which includes three sub-networks: the generator, the discriminator and the VGG pre-trained network. The pre-trained VGG network is used to extract feature maps of real and fake images. These feature maps are then used to calculate the VGG-based perceptual loss that compares two images in feature space.

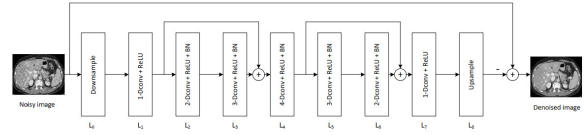


Fig. 5 The architecture of the proposed dilated CNN-based denoising model. The pre-processing layer L_0 and the post-processing L_8 are designed similarly as those in the FFDNet method [37]. Dilated convolution operators are used for all layers from L_1 to L_7 .

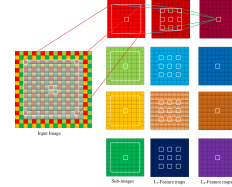


Fig. 6 Illustration of 3×3 2-Dconv operator at layer L_2 in the proposed architecture. 2-Dconv operator impacts on regions of 5×5 of feature maps at layer L_1 (pixels highlighted by white border), and thus on regions of size 7×7 and 14×14 in the four subimages and the input image, respectively.

Generally, the CNN-based methods significantly outperform traditional methods in both objective and subjective comparisons. How to effectively remove noise and artifacts while preserving subtle details is the biggest challenge in the problem of LDCT image reconstruction. For this purpose, the existing CNN-based methods focus on both the architecture of the network and the definition of the loss function.

3 The Proposed Denoising Method

The similarity of non-local pixels (usually computed based on patch comparison) plays a key role in the success of well-known image denoising methods, namely NLM [27] and BM3D [15]. In CNN, similar pixels often have the approximate values in the feature maps. Therefore, our idea is to extend receptive fields in deep CNN models. This helps similar non-local pixels in images better contribute to denoising.

3.1 Network Architecture

For the purpose of extending receptive fields, in the proposed network, the pre-processing and post-processing techniques proposed in [37] are embedded into a dilated residual convolutional neural network (DRN) using a combination of dilated convolutions (denoted by DConv) [49] and skip connections with the residual learning structure. The DRN architecture is inspired by SAR-DRN, proposed in [50]. The architecture of the proposed network is illustrated in Fig. 5. The proposed network consists of 9 layers, specifically: 1 pre-processing layer, followed by 7 non-linear mapping layers and then 1 post-processing one. The pre-processing layer, L_0 , performs down-sampling of the noisy input image into arranged sub-images. At this layer, an $n \times m$ image is decomposed into four sub-images with size of $\lceil \frac{n}{2} \rceil \times \lceil \frac{m}{2} \rceil$, as demonstrated in Fig. 6. These four sub-images are fed into the DRN consisting of seven 3×3 dilated convolution layers

(L_1, L_2, \dots, L_7) and two skip connections. In this work, the dilation factors of the 3×3 dilated convolutions from layer L_1 to layer L_7 are respectively set to 1, 2, 3, 4, 3, 2, and 1, similar to [50]. The two skip connections are employed to connect layer L_1 to layer L_3 and layer L_4 to layer L_7 . Unlike the original SAR-DRN network introduced in [50], in layers from L_2 to L_6 we use Dilated Convolution, Batch Normalization and ReLU. DRN (after layer L_7) generates 4 sub-images. These images then are up-sampled by the up-sample layer (L_8) to obtain an estimation of noise and artifacts in the input image (the same size as the input image). Finally, this estimate is combined with the input noisy image as inputs to residual learning, to obtain the denoised image.

The use of pre- and post-processing techniques and dilated convolution can strongly enlarge the receptive field while maintaining the kernel size. By using the dilated convolution with down-sampling, at layer L_1 the 1-Dconv operator with 3×3 filter can active a region of size 6×6 on the input image, two times larger compared to traditional convolution without down-sampling step. Fig. 6 describes the effect of the 2-Dconv operator at layer L_2 on the input image when using the dilated convolution with 3×3 kernel and factor 1. Visually, we can see that the 2-Dconv operator affects a region of size 14×14 on the input images. Therefore, large distance pixels can contribute for calculation of feature maps, that means the expansion of receptive fields leads to increase information for feature maps in the network. These advantages helps achieve effective denoising.

3.2 Loss Function

The loss function measures the difference between the predicted output (denoised LDCT images) and the ground-truth (NDCT images). While the architecture determines the complexity of the model, the loss function controls how to learn the denoising model from the training dataset. There exist many loss functions for image restoration in the literature. In this work, we use both the MSE loss and the VGG-based perceptual loss in the overall loss function. Per-pixel MSE loss often leads to fast convergence. However, the use of the MSE loss leads to over-smoothed edges and details in denoised images. Thus, the perceptual loss is applied to deal with these issues. Suppose the training set contains N image pairs $\{\mathbf{x}_i, \mathbf{y}_i\}_{i=1}^N$, and $\hat{\mathbf{y}}_i$ is the output of the CNN denoising model \mathcal{F} with input $\mathbf{x}_i, i = \{1, \dots, N\}$. The MSE loss is

$$\mathcal{L}_{\text{MSE}}(\Theta) = \frac{1}{2N} \sum_{i=1}^N \|\hat{\mathbf{y}}_i - \mathbf{y}_i\|^2 = \frac{1}{2N} \sum_{i=1}^N \|\mathcal{F}(\mathbf{x}_i, \Theta) - \mathbf{y}_i\|^2, \quad (3)$$

where Θ is a set of network parameters. The perceptual loss is determined in the feature space as follows:

$$\mathcal{L}_{\text{Perceptual}}(\Theta) = \frac{1}{2N} \left[\frac{1}{whd} \|\Phi(\mathcal{F}(\mathbf{x}; \Theta)) - \Phi(\mathbf{y})\|_F^2 \right], \quad (4)$$

where Φ is a feature extractor, w , h , and d are the width, the height and the depth of the feature space, respectively. In

this paper, we deploy the pre-trained VGG-19 network [51] for feature extraction.

To take advantage of the MSE loss and the perceptual loss, we used the linear combination of them as a loss function for the proposed model:

$$\mathcal{L}(\Theta) = \mathcal{L}_{\text{MSE}} + \lambda \mathcal{L}_{\text{Perceptual}}, \quad (5)$$

where λ is a non-negative weight, used for balancing the role of each loss function.

4 Experiments and Performance Evaluation

In this section, we present objective and subjective comparisons between the proposed method (named as DRN-LDCT) and the well-known BM3D method [52] as well as state-of-the-art CNN-based methods namely WGAN-VGG, RED-CNN, FD-VGG. Experiments were performed with the same dataset. To quantitatively evaluate the performance of the methods, we use three quality indices, namely PSNR (Peak Signal to Noise Ratio), SSIM (Structural Similarity), and FSIM (Feature Similarity). PSNR measures the intensity difference between denoised image and ground truth image (noise-free image). It cannot describe the subjective quality of the image, which are often very important in medical images. SSIM and FSIM better express the structure and feature similarity between the recovered image and the reference one, as compared to PSNR, and thus they better preserve important information of denoising methods.

4.1 Dataset

The ‘‘2016 Low Dose CT Grand Challenge’’ database supported by the National Institute of Health, the American Association of Physicist in Medicine, and Mayo Clinic [40], was used in this work for training the CNN models. This database contains 1mm and 3mm thickness CT slices of full-dose and simulated quarter acquired from 10 anonymous patients. In this works we only use 3mm CT images. We randomly selected 900 quarter-full dose image pairs from six patents to establish the training dataset. It consists of more than 600.000 patches pairs of size 64×64 randomly cropped at the same positions in low-dose and full-dose image pairs. For validation and testing, we randomly selected 300 image pairs from four remain patients. The full-dose CT images are considered as the ground-truths for computing the quality indices.

4.2 Parameter Setting

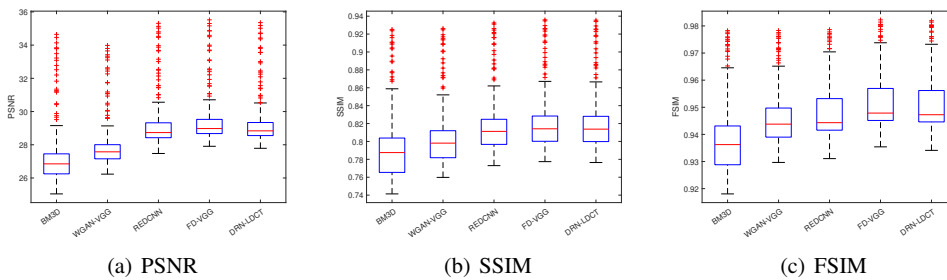
For the existing CNN-based models for LDCT image denoising (WGAN-VGG, RED-CNN, FD-VGG), parameters were set as guided by authors in the corresponding papers. For the proposed method, parameter λ in (5) was typically set to 0.1, the learning rate was set to 10^{-2} and decreased by half after every ten epochs. The number of epochs was set to 50. The Adam optimizer with default hyper-parameter values was used for training model parameters.

Table 1 Averaged PSNR, SSIM, and FSIM of different LDCT denoising methods, on testing dataset of 150 LDCT images.

Metric	BM3D	WGAN-VGG	RED-CNN	FD-VGG	DRN-LDCT (proposed)
PSNR	27.4802	28.1294	29.3713	29.5893	29.4554
SSIM	0.7963	0.8071	0.8198	0.8230	0.8226
FSIM	0.9397	0.9467	0.9480	0.9517	0.9511

Table 2 Network capacity and testing time.

Metrics	RED-CNN	WGAN-VGG	FD-VGG	DRN-LDCT (proposed)
Capacity (MB)	1.9	0.23	2	0.75
Testing time on GPU (s)	0.0014	0.001	0.0045	0.0047
Testing time on CPU (s)	4.4	0.13	0.15	0.09

**Fig. 7** Performance of BM3D, WGAN-VGG, RED-CNN, FD-VGG and DRN-LDCT.

4.3 Objective and Subjective Comparison

150 LDCT images from the testing dataset were used to evaluate the performance of the methods. Quality indices are computed based on comparing denoised LDCT images and their associating full-dose images. Experimental results are shown in Table 1 and Fig. 7. Specifically, Table 1 shows the average values of PSNR, SSIM, and FSIM computed on the testing dataset. Figs. 7(a)-7(c) represent basic statistical measurements of PSNR, SSIM and FSIM values of the methods over 150 samples of the testing dataset, respectively. As can be seen, the quality indices of the proposed method (DRN-LDCT) were comparable to those of RED-CNN and FD-VGG, while significantly higher than those of BM3D and WGAN-VGG. To further illustrate the effectiveness of the proposed method, Fig. 8 shows denoising results of different methods performed on an LDCT image of the abdominal region. This image contains two liver lesions which are highlighted by two small rectangles. Visually, as compared to BM3D and WGAN-VGG, noise was better suppressed by RED-CNN and FD-VGG. Small dark signals in highlighted regions were well preserved by DRN-LDCT and by the other CNN-based methods.

Another example is shown in Fig. 9. In this figure, the noisy image for testing is an abdominal LDCT image of another patient. The testing image has two liver lesions as marked by two small rectangles [40] (see Fig. 9(a)). By subjective comparison of the denoised images by the methods

and the NDCT images, we can see that globally the denoising effectiveness of DRN-LDCT was equivalent to that of RED-CNN and FD-VGG, while being slightly better than BM3D and WGAN-VGGs. In the regions of interest, the proposed method seemed to outperform the other methods in preserving the structures in the larger lesion. Moreover, the small dark point in the small highlighted region in the image denoised by the proposed method was also clearer.

4.4 Network Capacity and Computational Time

Table 2 shows the capacity and the average computational time of the different CNN-based methods under comparison. The computational time was recorded from experiments performed with testing images of size 512×512 , on a CPU Intel(R) Core(TM) i9-9900K CPU @ 3.60GHz $\times 8$ as well as on the GeForce RTX 3090 GPU card. As can be seen in Table 2, the capacity of DRN-LDCT was 0.75 MB, which is significantly smaller than that of RED-CNN (1.9 MB) and FD-VGG (2.0 MB), while the denoising effect was nearly equivalent. The computational time on CPU of DRN-LDCT was the lowest.

5 Conclusions

This paper has presented a brief overview of learning-based methods for denoising of LDCT images, from simple patch-based learning methods to modern CNN-based methods. In addition, we have proposed a very competitive denoising

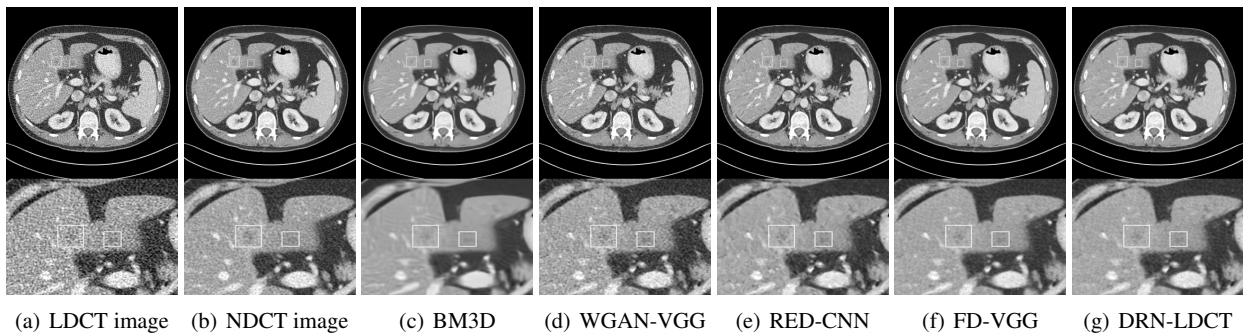


Fig. 8 Denoising results of an abdominal LDCT image. (a) noisy LDCT image in which the rectangles illustrate two desired regions of interest, (b) NDCT image corresponding to the LDCT image, (c)-(g) images denoised by BM3D, WGAN-VGG, RED-CNN, FD-VGG and DRN-LDCT.

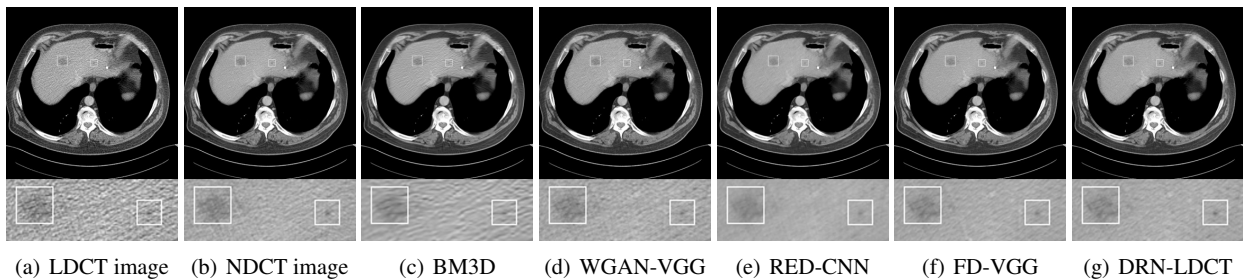


Fig. 9 Denoising results on another LDCT image of the abdomen.

method for LDCT images, namely DRN-LDCT, which is based on the idea of extending the receptive field in an end-to-end CNN architecture. The proposed denoiser is lighter as compared to some state-of-the-art methods such as RED-CNN and FD-VGG, while the quality was equivalent. The experimental results have demonstrated the effectiveness of our method over several leading state-of-the-art LDCT denoising methods. Future work may look into finding more appropriate loss functions and try to insert the attention mechanism in the network architecture in order to focus more useful features in medical images.

References

1. G. N. Hounsfield, "Computerized transverse axial scanning (tomography): Part 1. description of system," *The British J. Radiology* **46**(552), 1016–1022 (1973).
2. J. Wang *et al.*, "Sinogram noise reduction for low-dose CT by statistics-based nonlinear filters," in *Med. Imag. 2005: Image Processing*, **5747**, 2058–2066 (2005).
3. J. Wang *et al.*, "Penalized weighted least-squares approach to sinogram noise reduction and image reconstruction for low-dose X-ray computed tomography," *IEEE Trans. Med. Imag.* **25**(10), 1272–1283 (2006).
4. A. K. Hara *et al.*, "Iterative reconstruction technique for reducing body radiation dose at CT: feasibility study," *Amer. J. Roentgenology* **193**(3), 764–771 (2009).
5. M. Beister *et al.*, "Iterative reconstruction methods in X-ray CT," *Physica medica* **28**(2), 94–108 (2012).
6. W. Stiller, "Basics of iterative reconstruction methods in computed tomography: A vendor-independent overview," *European J. Radiol.* **109**, 147–154 (2018).
7. A. Buades *et al.*, "A review of image denoising algorithms, with a new one," *Multiscale Modeling & Simulation* **4**(2), 490–530 (2005).
8. Z. Li *et al.*, "Adaptive nonlocal means filtering based on local noise level for CT denoising," *Med. Phys.* **41**(1), 011908 (2014).
9. M. Green *et al.*, "Efficient low-dose CT denoising by locally-consistent non-local means (LC-NLM)," in *Proceedings MICCAI 2016*, Part III, 423–431 (2016).
10. M. Elad and M. Aharon, "Image denoising via sparse and redundant representations over learned dictionaries," *IEEE Trans. Image Process.* **15**(2) (2006).
11. Y. Chen *et al.*, "Improving abdomen tumor low-dose CT images using a fast dictionary learning based processing," *Phys. Med. Biol.* **58**(16), 5803 (2013).
12. Y. Chen *et al.*, "Artifact suppressed dictionary learning for low-dose CT image processing," *IEEE Trans. Med. Imag.* **33**(12), 2271–2292 (2014).
13. A. Pizurica *et al.*, "A versatile wavelet domain noise filtration technique for medical imaging," *IEEE Trans. Med. Imag.* **22**, 323–331 (2003).
14. J. Portilla *et al.*, "Image denoising using scale mixtures of Gaussians in the wavelet domain," *IEEE Trans. Image Process.* **12**(11), 1338–1351 (2003).

15. K. Dabov *et al.*, "Image denoising by sparse 3-D transform-domain collaborative filtering," *IEEE Trans. Image Process.* **16**(8), 2080–2095 (2007).
16. H. Lu *et al.*, "Noise properties of low-dose CT projections and noise treatment by scale transformations," in *Nuclear Sci. Symp.*, **3**, 1662–1666 (2001).
17. D.-H. Trinh *et al.*, "Novel example-based method for super-resolution and denoising of medical images," *IEEE Trans. Image Process.* **23**(4), 1882–1895 (2014).
18. H. Chen *et al.*, "Low-dose CT with a residual encoder-decoder convolutional neural network," *IEEE Trans. Med. Imag.* **36**(12), 2524–2535 (2017).
19. Q. Yang *et al.*, "Low dose CT image denoising using a generative adversarial network with Wasserstein distance and perceptual loss," *IEEE Tr. Med. Imag.* (2018).
20. X. Yi and P. Babyn, "Sharpness-aware low-dose CT denoising using conditional generative adversarial network," *J. Dig. Imag.* **31**(5), 655–669 (2018).
21. C. You *et al.*, "Structurally-sensitive multi-scale deep neural network for low-dose CT denoising," *IEEE Access* **6**, 41839–41855 (2018).
22. N. T. Trung *et al.*, "Robust denoising of low-dose CT images using convolutional neural networks," in *Proc. NICS*, 506–511, (2019).
23. D. H. Trinh *et al.*, "Medical image denoising using kernel ridge regression," in *Proc. ICIP*, 1597–1600, (2011).
24. J. Ma *et al.*, "Low-dose computed tomography image restoration using previous normal-dose scan," *Med. Phys.* **38**(10), 5713–5731 (2011).
25. D. H. Trinh *et al.*, "An optimal weight method for CT image denoising," *J. Electro. Sci. Tech.* **10**(2) (2012).
26. J. Ma *et al.*, "Low-dose computed tomography image restoration using previous normal-dose scan," *Med. Phys.* **38**(10), 5713–5731 (2011).
27. C. Kervrann and J. Boulanger, "Optimal spatial adaptation for patchbased image denoising," *IEEE Trans. Image Process.* **15**, 2866–2878 (2006).
28. T.-T. Nguyen *et al.*, "An efficient example-based method for CT image denoising based on frequency decomposition and sparse representation," in *Proc. ATC*, 293–296, (2016).
29. D. Zoran and Y. Weiss, "From learning models of natural image patches to whole image restoration," in *Proc. ICCV*, 479–486 (2011).
30. H. C. Burger *et al.*, "Learning how to combine internal and external denoising methods," in *Proc. GCPR*, **8142**, 121–130 (2013).
31. E. Luo *et al.*, "Adaptive image denoising by targeted databases," *IEEE Trans. Image Process.* **24**(7), (2015).
32. V. Jain and H. S. Seung, "Natural image denoising with convolutional networks," in *NIPS*, 769–776 (2008).
33. D. Li, "Support vector regression based image denoising," *Image Vis. Comput.* **27**(6), 623–627 (2009).
34. H. C. Burger *et al.*, "Image denoising: Can plain neural networks compete with BM3D?," in *Proc. CVPR*, 2392–2399 (2012).
35. D. H. Trinh *et al.*, "An effective example-based denoising method for CT images using markov random field," in *Proc. ATC*, 355–359 (2014).
36. K. Zhang *et al.*, "Beyond a gaussian denoiser: Residual learning of deep CNN for image denoising," *IEEE Trans. Image Process.* **26**(7), 3142–3155 (2017).
37. K. Zhang *et al.*, "FFDnet: Toward a fast and flexible solution for CNN-based image denoising," *IEEE Trans. Image Process.* **27**(9), 4608–4622 (2018).
38. Z. Yue *et al.*, "Variational denoising network: Toward blind noise modeling and removal," in *Proc. NeurIPS*, 1688–1699 (2019).
39. S. Anwar and N. Barnes, "Real image denoising with feature attention," in *Proc. ICCV*, 3155–3164, (2019).
40. NIH, AAPM, and Mayo Clinic, "Low dose CT grand challenge," (2016).
41. J. Leuschner *et al.*, "LoDoPaB-CT, a benchmark dataset for low-dose computed tomography reconstruction," *Scientific Data* **8**(109) (2021).
42. X.-J. Mao *et al.*, "Image restoration using very deep convolutional encoder-decoder networks with symmetric skip connections," *arXiv:1603.09056* (2016).
43. O. Ronneberger *et al.*, "U-net: Convolutional networks for biomedical image segmentation," in *Proc. MICCAI*, **9351**, 234–241 (2015).
44. M. Li *et al.*, "SACNN: self-attention convolutional neural network for low-dose CT denoising with self-supervised perceptual loss network," *IEEE Trans. Med. Imag.* **39**(7), 2289–2301 (2020).
45. H. Shan *et al.*, "Correction for "3D convolutional encoder-decoder network for low-dose CT via transfer learning from a 2D trained network"," *IEEE Trans. Med. Imag.* **37**(12), 2750 (2018).
46. I. J. Goodfellow *et al.*, "Generative adversarial networks," *arXiv preprint arXiv:1406.2661* (2014).
47. M. Arjovsky *et al.*, "Wasserstein GAN," *arXiv:1701.07875* (2017).
48. W. Du *et al.*, "Visual attention network for low-dose CT," *IEEE Sig. Process. Let.* **26**(8), 1152–1156 (2019).
49. F. Yu and V. Koltun, "Multi-scale context aggregation by dilated convolutions," in *Proc. ICLR*, (2016).
50. Q. Zhang *et al.*, "Learning a dilated residual network for sar image despeckling," *Rem. Sens.* **10**(2), 196 (2018).
51. K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," in *Proc. ICLR*, (2015).
52. D. Kang *et al.*, "Image denoising of low-radiation dose coronary CT angiography by an adaptive block-matching 3D algorithm," in *Med. Imag. 2013: Image Process*, 86692G (2013).