

Fully Convolutional Architecture for Low-Dose CT Image Noise Reduction

S Badretale¹, F Shaker², P Babyn³ and J Alirezaie⁴

^{1,4} Department of Electrical and Computer Engineering, Ryerson University, Toronto, ON, M5B 2K3 Canada

² Department of AI, Faculty of Computer Engineering, University of Isfahan, Isfahan, 81746, Iran

³ Department of Medical Imaging, University of Saskatoon, Saskatoon, SK, S7N 0W8, Canada

¹sbadretale@ryerson.ca, ²f.shaker@eng.ui.ac.ir,
³paul.babyn@saskatoonhealthregion.ca, ⁴javad@ryerson.ca

Abstract. One of the critical topics in medical low-dose Computed Tomography (CT) imaging is how best to maintain image quality. As the quality of images decreases with lowering the X-ray radiation dose, improving image quality is extremely important and challenging. We have proposed a novel approach to denoise low-dose CT images. Our algorithm learns directly from an end-to-end mapping from the low-dose Computed Tomography images for denoising the normal-dose CT images. Our method is based on a deep convolutional neural network with rectified linear units. By learning various low-level to high-level features from a low-dose image the proposed algorithm is capable of creating a high-quality denoised image. We demonstrate the superiority of our technique by comparing the results with two other state-of-the-art methods in terms of the peak signal to noise ratio, root mean square error, and a structural similarity index.

1. Introduction

Computed Tomography (CT) imaging is widely used in diagnosing various medical conditions. However, the amount of x-ray needed to produce a high-quality CT image is relatively high which may result in an increasing risk of developing cancer. Therefore, it is desirable to decrease the X-ray dose for CT imaging. However, low dose CT images are highly noisy which makes them unreliable for diagnostic purposes. There have been many studies aiming to remove the noise from the low dose CT images without affecting the details of underlying image structures. X-ray CT images are reconstructed using special algorithms from the X-ray projection data. The main source of noise in X-ray projection data is the statistical fluctuations of X-ray quanta reaching the detectors. This noise follows the Poisson distribution. However, after the application of reconstruction algorithms and image processing steps to the projection data, the distribution of the noise in the resulting CT image is unknown and therefore is hard to model.

There are several approaches for denoising CT images; one is to use the original sinogram data in which the noise distribution is known to be Poisson, the other approach is to remove the noise from CT images in image domain after their reconstruction in which the noise distribution is unknown. In



this research, we focus on the second approach since accessing the raw projection data is not always possible.

There has been a number of works regarding the denoising of CT images in the image domain. Ghadrdan et al. used dictionary learning combined with wavelet transform [1]. They used wavelets to extract the most suitable features in images to obtain accurate dictionary atoms for denoising algorithm. To improve the computational complexity, the number of clusters was decreased. They compared their method with the clustering based sparse representation (CSR) and the K-SVD method. Their method shows some improvement in computational complexity and noise reduction over the compared methods. Denoising of CT images using dictionary learning was also studied by Khodabandeh et al. [2].

In another study by Hashemi et al. [3], a noise confidence region evaluation (NCRE) method was used in combination with the block matching and 3D filtering (BM3D) method to adaptively estimate the noise properties instead of using an independent identical additive Gaussian noise model. The NCRE adaptively updates the regularization parameters at the end of each iteration by validating the result of that iteration.

Deep learning is an emerging area in AI with many successful applications in recent years. Unlike traditional multilayer NNs, which are fully connected, deep neural networks have sparse connections. Convolutional neural networks (CNN) are a subset of deep neural networks. The problem of denoising can be considered as a designing task in which an appropriate filter is constructed to remove the noise. This view aligns well with CNN in which a combination of filters in different layers is learned through real examples. The nonlinear property of the overall filter lends itself to complex distributions, which otherwise are hard to design.

Using the methods that rely on learning will be more appropriate in this scenario since the noise model is learned from the actual data rather than by estimating the complicated noise model. In this study, deep learning is used in denoising of the CT images.

In section 2, the details of our algorithm are presented. Section 3 contains the results and comparisons of our techniques and finally, in section 4 some concluding remarks are provided.

2. Materials and Methods

2.1. Problem Formulation

In low-dose CT imaging, the noise makes the edges and fine image features blurred and distorted. The goal of low-dose CT image denoising is to map the noisy and distorted image onto the clean image such that the noise is reduced or fully removed. In our research, we have considered actual low-dose and the corresponding normal-dose images, and the interfering normal-dose CT images by Gaussian or Poisson noise are not considered, so the noise is not modeled in this work.

Let X be the desired normal-dose CT image and Y be the low-dose counterpart. The Denoising function is defined as F .

$$F = \operatorname{argmin}_F \|F(Y) - X\|_2^2 \quad (1)$$

The objective here is to determine the function that creates the closest image to X from noisy image Y . Function F can be considered as a nonlinear filter whose coefficients are not known but can be learned through training images. In this study, convolutional neural networks are used to learn such a filter.

2.2. Overall Framework

The denoising convolutional neural network is a modified CNN whose its parameters, e.g. weights and biases should be learned and calculated with the backpropagation training. An overview of the framework is shown in Figure 1.

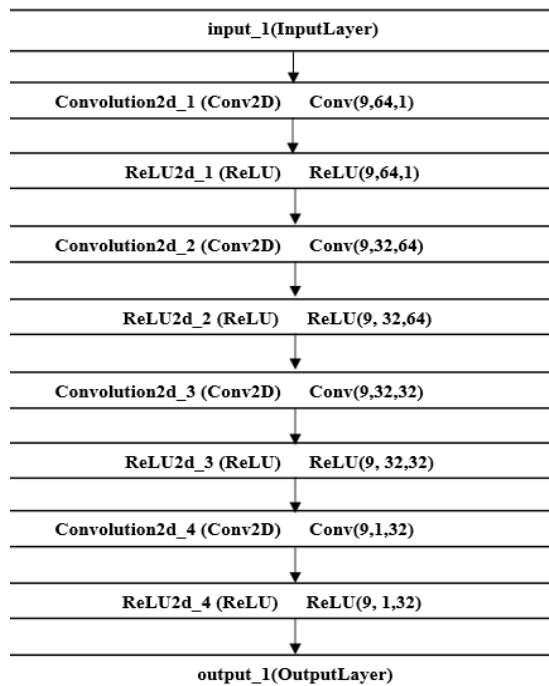


Figure 1. A network structure of the end-to-end nonlinear mapping denoising convolutional neural network.

The network that we have utilized in this study consists of four layers. Each of these layers consists of a number of convolution filters followed by an activation layer. The loss function is calculated at the end of every feed-forward step to update the parameters (weights and biases). The number of feature maps that are used from the first to the last layer is 64, 32, 32, and 1 respectively. The size of all convolutional filters is 9×9 . The input of the network is the low-dose noisy image (Y), that passes through some convolutional layers to extract sets of feature maps to produce a denoised image $F(Y)$. Hence, a number of feature maps, which require going through the activation function, form the first convolutional layer. The subsequent convolutional layers receive the output of the previous layer then pass its results through activation functions for nonlinear processing to make higher-level feature maps in order to create a corresponding noise free image. As a result, the network trained hierarchically structured feature maps from low-level (blobs, edges, etc.) to high-level (more complex and detailed shapes). Moreover, we came up with a pre-train convolutional network, which can be used as a transfer learning.

2.3. ReLU Nonlinearity

Nonlinear processing as an activation function accompanied by the pooling layer makes crucial improvements to feature extraction and makes the network rotation, shift, and scale invariant. Apart from the traditional logistic sigmoid function as an activation function, a number of nonlinear rectifying functions were introduced including the absolute positive function. Among all the activation functions, we decided to choose Rectified Linear Unit (ReLU), since several studies on deep convolutional neural networks have indicated discernible development by applying rectified nonlinear processing, and non-saturating nonlinear functions are greatly accelerated compared to saturating activation functions. By definition, the Rectified Linear Unit (ReLU) model for every input x is

$$G(x) = \text{Max}(0, x) \quad (2)$$

Mainly due to the non-saturating and linear form of ReLU, the convergence of the stochastic gradient descent is much faster than the hyperbolic tangent (tanh) and Sigmoid functions. Therefore, the training time becomes noticeably shortened. Furthermore, it does not have any reciprocal and exponential computation. That is just a simple thresholding so its implementation is much less expensive than other available alternatives. In addition, many studies have indicated that to acquire a high convergence rate RLU could be applied as an activation function [4]. We have detailed our

results in Figure 2, indicating that by applying ReLU a greater PSNR can be achieved. Besides, in order to get a low error (RMSE) of 0.02 in the validation set, it needs much less number of backpropagations, 9×10^8 for ReLU as 1.6×10^8 for tanh function. The dataset consists of 500 low-dose CT images and the corresponding normal-dose CT images were taken from a patient CT chest.

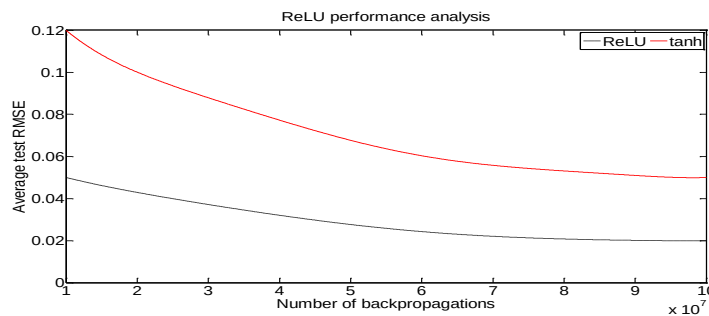


Figure 2. Comparison of tanh and ReLU as an activation function.

2.4. Implementation Details

The proposed deep convolutional neural network can be implemented by using a large set of low-dose and the corresponding normal-dose CT images. Due to a lack of sample data, low-dose CT images have been created inaccurately by adding Poisson or Gaussian noise to the normal-dose CT images. We implemented our method based on real low-dose and normal-dose datasets. We adopt widely known peak signal-to-noise ratio (PSNR), root mean squared error (RMSE), and Structural similarity (SSIM) [5] as quantitative evaluation metrics which express high correlation with the human perceptual scores to compare the result $F(Y)$ to the normal-dose image (X). Weights and biases, which are the network parameters are updated by standard stochastic gradient descent via backpropagation.

We use two different CT datasets. One dataset has 500 axial chest CT images obtained from a clinical patient. Images are from the same imaging session and it consists of two group images. One group image is taken using the higher dose of X-ray (exposure 200 mAs) and the other one is taken with a lower dose (exposure 25 mAs). Therefore, there is a one to one correspondence between images. Also, the phantom size and slice thickness is fixed. The second dataset has 500 standard phantom CT images, same number of low-dose (25 mAs) and normal-dose (200 mAs) images, containing spheres with different spacing lines with a wide variety of contrast [6]. It should be noted that creating normal-dose CT image by adding Poisson noise is not accurate. This makes the denoising algorithms impractical since the noise distribution in real low-dose CT images is not known. All the input images are in Digital Imaging and Communications in Medicine (DICOM) format.

The output images are also saved in DICOM format. The only pre-processing step is the normalization of all images between 0 and 1. Furthermore, we randomly cropped the training, validation and tested sets as 33×33 pixels sub-images with the stride of 14. Therefore, we obtain several thousand created small sub-images rather than overlapping patches and we do not input the original size images through the network.

The biases are initialized to zero. The filter weights of each layer for the deep convolutional neural network are initialized by drawing randomly from a Gaussian distribution with zero mean and standard deviation 0.001. Furthermore, the learning rate for the first three layers are initially set to 0.01 and is going to decrease by a factor of 0.9 every 10000 iterations. For the last convolution layer, the learning rate is initialized to 0.001 with the same decaying factor as the other layers, as well as the momentum term of 0.9.

3. Experimental Results

The proposed algorithm was validated by experiments using phantom data and real in vivo CT dataset and the results are compared with BM3D [7] and Simultaneous Sparse Coding (SSC-GSM) [8] which is claimed as the state-of-the-art image restoration by the authors.

We randomly shuffle images in each dataset in order to reflect enough diversity in the training, validation, and test data. The size of images is 512×512 . The training, validation, and test data include 50%, 20%, and 30% of each dataset respectively.

The proposed algorithm is implemented using the Caffe package [9] and Matlab R2016a. The processor is Intel core i7 CPU 3.4GHz and 16GB memory and the Graphics Card is GeForce GTX 1070. The training time is about 10 hours and the test time was less than a second for each image.

Table 1. indicates the average denoising results for the proposed deep convolutional neural network algorithm, BM3D, and SSC-GSM for the real CT dataset. As it is obvious in Tables 1. the proposed deep learning method provides the highest scores in all quantitative metrics. It is worth noting that the result of our proposed method is based on the output from 10^8 backpropagation. Thus, it has the best performance amongst all the mentioned methods.

Table 1. Comparison of performance on the chest CT dataset.

Eval. Mat	Proposed Algorithm	SSC-GSM	BM3D
PSNR	32.112	28.754	28.298
SSIM	0.834	0.786	0.736
RMSE	0.198	0.376	0.389

Figure 3. Visualizes performance of sample output images for different methods to grasp more intuitive feeling. All the images are drawn randomly. The first two rows are from the patient CT dataset. The last row is from the phantom CT dataset. As can be observed, BM3D and SSC-GSM perform better on the phantom data due to its structured pattern of the spheres. However, our proposed method still produces better images as it removes more artifacts and creates sharper edges. That is to say, our results on the patient and phantom datasets demonstrate that clinically important details are better preserved (higher SSIM) and the noise is significantly reduced (Higher PSNR).

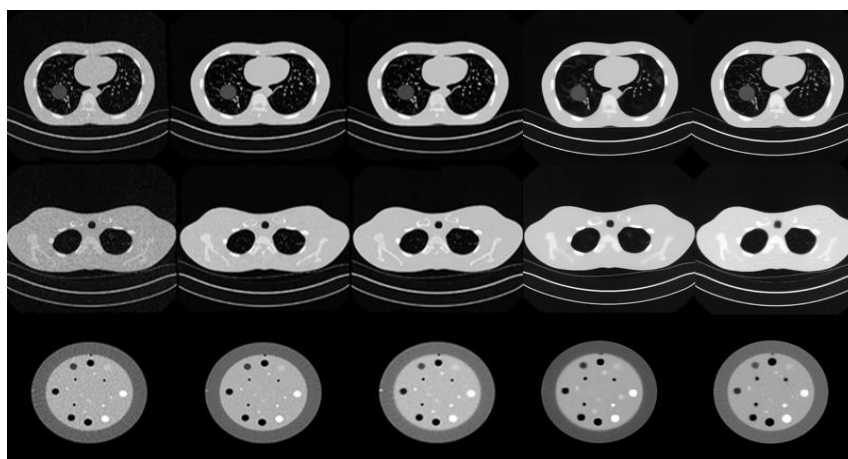


Figure 3. Visual comparison of denoising outputs. From left to right: Low-dose, Normal-dose, Proposed algorithm, BM3D method, SSC-GSM method.

In our method, the PSNR is 32.434, 34.761, and 44.234 from top to bottom. In BM3D method, the PSNR is 29.238, 31.1, and 41.376. And in SSC-GSM algorithm, the PSNR is 30.345, 32.964, and 42.76. The result of the denoising methods over low-dose CT image is shown in Figure 4.

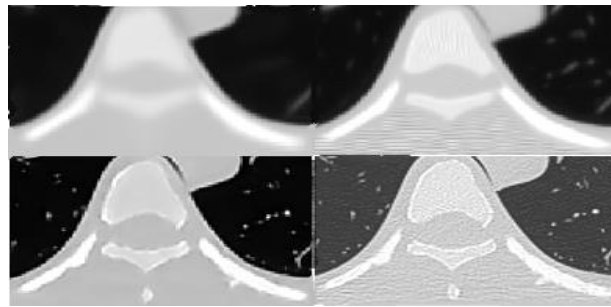


Figure 4. Zoomed images of the result of the denoising methods. Top left: BM3D, Top right: SSC-GSM, Bottom left: Proposed method, Bottom right: normal-dose.

4. Conclusion and Future work

We have introduced a novel deep learning convolutional neural network algorithm for denoising low-dose CT images. The proposed framework has a generalized formulation capable of handling a wide range of CT images with arbitrary size. We used an efficient end-to-end mapping between low-dose and normal-dose CT images. We demonstrated that the proposed algorithm has an outstanding merit due to its simplicity and surpassing of the performance of the most common methods.

We are further analyzing the effects of different strategies such as the number of convolution layers and kernel size. Also, higher quality CT images might be obtained by dataset expansion and data augmenting.

References

- [1] Ghadrdan *et al.* Low-dose Computed Tomography Image Denoising based on Joint Wavelet and Sparse representation. *Engineering in Medicine and Biology Society (EMBC). 2014. 36th Annual International Conference of the IEEE*, 3325–3328.
- [2] Khodabandeh A, Alirezaie J, Babyn P, Ahmadian A. Computed Tomography Image Denoising by Learning to Separate Morphological Diversity. 2015. *Telecommunications and Signal Processing (TSP). 38th International Conference on*: 513–517.
- [3] Hashemi S, Paul N, Beheshti S, Cobbold R. 2015. Adaptively Tuned Iterative Low Dose CT Image Denoising. 2015, 1: 1–14.
- [4] Arora R, Basu A, Mianjy P, Mukherjee A, Understanding Deep Neural Networks with Rectified Linear Units. 2016. arXiv preprint arXiv:1611.01491.
- [5] Wang Z, Bovik C, Sheikh H, Simoncelli E, Image quality assessment: from error visibility to structural similarity. 2004. *Image Processing, IEEE Transactions on* 13, 4: 600–612.
- [6] Gavrielides M, Kinnard L, Myers K, Peregoy J, Pritchard W, 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, 2010. *Optics Express*, vol. 18, n.14, pp. 15244-15255.
- [7] Dabov K, Foi A, Egiazarian K, Video denoising by sparse 3D transform-domain collaborative filtering. 2007. *European Signal Processing Conference* 16, 8, <https://doi.org/10.1109/TIP.2007.901238>, 145–149.
- [8] Dong W, Shi G, Ma Y, Li X, Image Restoration via Simultaneous Sparse Coding: Where Structured Sparsity Meets Gaussian Scale Mixture. 2015. *International Journal of Computer Vision* 114, 2. <https://doi.org/10.1007/s11263-015-0808-y>, 217-232.
- [9] Jia Y, Shelhamer E, Donahue J, Karayev S, Long J, Girshick R, Guadarrama S, Darrell T. Caffe: Convolutional Architecture for Fast Feature Embedding. 2014. *ACM International Conference on Multimedia*: <https://doi.org/10.1145/2647868.2654889>, 675–678.