Few-View CT Reconstruction Method Based on Deep Learning

Ji Zhao, Zhiqiang Chen, Li Zhang, Xin Jin

Abstract—**To reduce patient's dose, few-view CT reconstruction promises to be a good attempt. The key to better reconstruction is the sparse view artifacts. In recent years, DL(deep learing) has attracted a lot of attention because its outstanding performance in image processing. We propose a deep learning method for few-view CT reconstuction. Our method directly learns an end-to-end mapping between the full-view/few-view reconstruction. The mapping is represented as a deep convolutional neural network (CNN) that takes the few-view reconstruction image as the input and outputs the full-view one. We further show that traditional Dictionary Learning based reconstruction methods can also be viewed as a deep convolutional network. But unlike traditional methods that handle each component separately, our method jointly optimizes all layers. Our deep CNN has a lightweight structure, yet demonstrates state-of-the-art reconstruction quality, and achieves fast speed for practical on-line usage. We explore different network structures and parameter settings to achieve trade-offs between performance and speed.**

Index Terms—**Computed Tomography, few-view, Deep Learning, CNN**

I. Introduction

CT has been widely used in clinical diagnosing process, and rapid growing radiation dose [1] can be harmful to people health [2] , which has attracted a lot of concerns, such as in UK [3] . To reduce patient's dose, few-view CT reconstruction promises to be a good attempt, The key problem is that using conventional few-view reconstruction may bring artifacts into reconstruction image, because few-view reconstruction is inherently ill-posed since a multiplicity of solutions exist for any given few-view CT projection data. In other words, it is an underdetermined inverse problem, of which solution is not unique.

By the time, we have two ways to reconstruct images in ill-posed problem. The first one is processing the projection, such as inpainting or filtering. La Riviere proposed smoothing the sinogram by a penalized likelihood technique [4] . Wang et al. used a penalized weighted least-squares regularizer to get a less noisy sinogram [5] . The other way is the iterative reconstruction (ie. ordered-subsets convex algorithm [6]).

Thanks to compressed sensing theory, which allows us to restore sparse signals from less data(less than Shannon Nyquist sampling theorem requires) [7] [8] , we can reconstruct image from incomplete data. Because in common cases, reconstruction images are not sparse at all, we need to find a domain in which the signal can be represented sparsely. So we can make the attempt to remove the sparse view artifacts is typically mitigated by constraining the solution space by prior information of sparsity.

One common assumptions is the discrete gradient transform (DGT) is sparse usually, which leads to total variation (TV) regularization. For its simplicity, TV is the most popular regularizer in state-of-the-art works [9] ,including few-view, limited-angle, and interior problems [10]. And there are many works to enhence TV method. Tian Zhen et al. proposed EPTV(edge-preserving total variation) to perform low-dose reconstruction. [11] Xin Jin et al. proposed ATV(anisotropic TV) for limited-angle CT reconstruction. [12].

However, the TV method is still limited. First, the TV regularizer is global, so it cannot directly reflect structures of an object. Second, small structures and noise can be confused in the TV minimization step. For the reasons above, images reconstruction with TV regularizer can have a blocky appearance because of noisy cases and can lose some small features in incomplete or noisy cases. To overcome the shortcoming, learning based method is proposed. Dictionary learning method is learning-based method,a dictionary is an over-complete basis, which is learned from application-specific image training sets. Qiong Xu et al. has adopted dictionary learning methods in CT reconstruction, which has [13], [14] The dictionary learning based method is thought to preform better than any generic sparse transform in application-specificsense as the dictionary is learned from training images.

The double-dictionary-based method is one of the representative few-view reconstruction methods. This method involves several steps in its solution pipeline. First, overlapping patches are densely cropped from the input image and pre-processed (e.g.,subtracting mean and normalization). These patches are then encoded by a dictionary, trained from few-view reconstruction. The sparse coefficients are passed into a high-resolution dictionary for reconstructing high-resolution patches. Reconstructed patches are aggregated (e.g., by weighted averaging) to produce the final output. This pipeline is shared by many methods, which pay particular attention to learning and optimizing the dictionaries or building efficient mapping functions.

However, the rest of the steps in the pipeline have been rarely optimized or considered in an unified optimization framework. In this paper, we show that the aforementioned pipeline is equivalent to a deep convolutional neural network.

This work was supported by grants from the National Natural Science Foundation of China (No. 11525521 and No. 61527807).

Ji Zhao, Zhiqiang Chen and Li Zhang are with Key Laboratory of Particle and Radiation Imaging (Tsinghua University), Ministry of Education, Beijing 100084, China and the Department of Engineering Physics, Tsinghua University, Beijing 100084, China. Xin Jin is with NUCTECH Company limited, Beijing 100084, China. (Corresponding author: Zhiqiang Chen, czq@tsinghua.edu.cn)

978-1-5090-1642-6/16/$31.00 ©2016 IEEE

Motivated by this fact, we consider a convolutional neural network (CNN) that directly learns an end-to-end mapping between few-view and full-view reconstructed images.

Recently, we proposed a deep learning based reconstruction method to address the low-dose CT reconstruction problem. Our method consists of two steps. The first step is the iterative reconstruction process that satisfying the data constraint and the second step is the CNN. whose role is similar to dictionary-based sparsification or the TV minimization Our method differs fundamentally from existing double-dictionary-learning method, because ours isn't based on the patch space. We use hidden layers to achieve these effects. Furthermore, we can also use convolutional layers for the patch extraction and aggregation steps, so they can be involved in the optimization. In our method, the entire pipeline is fully obtained through learning.

The proposed method has several appealing properties. First, we design the structure for simplicity and the layers in the network have clear physics meaning. Second, our method is faster than many state-of-art dictionary-learning methods, because it is fully feed-forward and does not require solving optimization problem in use. Our method is able to achieve fast speed for practical usage even on CPUs. Third, the network trained from larger and more diverse datasets can improve the quality of the reconstruction image proven by experiments. While larger datasets/models can present challenges for existing example-based methods.

The rest of this paper is organized as follows. In Section II, we will introduce the theory and methods of CNN. In Section III, we will report representative results for few-view projections, and compare the performance of our proposed methodology. Finally, in Section V we will do some discussion and conclude the paper.

II. Theory and Methods

A. Network and Formulation

Let us consider the few-view CT reconstuction image, denoted as $\mathbf{Y}$. Our goal is to recover from $\mathbf{Y}$ an image $F(\mathbf{Y})$ that is as similar as possible to the ground truth full-view image $\mathbf{X}$. We wish to learn a mapping F , which conceptually consists of three operations: patching, mapping and restoring.

We can see that all these operations form a convolutional neural network. An overview of the network is depicted in Figure 1. Next we detail our definition of each operation.

1) Patch extraction and representation: This operation extracts (overlapping) patches from the init image $\mathbf{Y}$ and represents each patch as a high-dimensional vector.

Current dictionary-learning methods represent them by a set of pre-trained bases trained as dictionary item. This is equivalent to convolving the image by a set of filters, each of which is a basis. In our method, we involve the optimization of these bases into the optimization of the network. Formally, our first layer is expressed as an operation F_1 :

$$F_1(\mathbf{Y}) = \max(0, W_1 * Y + B_1) \tag{1}$$

where W_1 and B_1 represent the filters and biases respectively, and $*$ denotes the convolution operation. These result vectors comprise a set of feature maps, of which the number equals to the dimensionality of the vectors.

2) Non-linear mapping: This operation nonlinearly maps each high-dimensional vector onto another high-dimensional vector. These vectors comprise another set of feature maps, which we wish to contain the information from the ground truth full-view image.

We use 3×3 or 5×5 patches of the feature map to determine the pixel non-linearly in output vector. The operation is

$$F_2(\mathbf{Y}) = \max(0, W_2 * F_1(Y) + B_2) \tag{2}$$

It is possible to add more convolutional layers to increase the non-linearity.

3) Restoration: This operation aggregates the above patch-wise representations to generate the final reconstution image. This image is expected to be similar to the ground truth X.

In the traditional methods, the patches are often averaged to produce the final full image. The averaging can be considered as an average filter on a set of feature maps. So we use the filter to restore the image:

$$F(\mathbf{Y}) = \max(0, W_3 * F_3(Y) + B_3) \tag{3}$$

B. Relationship to Dictionary-Learning-Based Methods

We show that the dictionary-learning-based SR methods can be viewed as a convolutional neural network. Figure 3 shows an illustration.

In the dictionary-learning-based methods, let us consider that an $f_1 \times f_1$ patch is extracted from the input image. Then the methods will first project the patch onto a dictionary, equivalent to applying n_1 linear filters on the input image. This is illustrated as the left part of Figure 2.

The solver will then iteratively process the coefficients and the outputs of this solver are another set of coefficients. These output coefficients are the representation of another dictionary.(e.g., see the Feature-Sign solver [15]) See the middle part of Figure 2. However, the sparse coding solver is not feed-forward, i.e.,it is an iterative algorithm. On the contrary, our non-linear operator is fully feed-forward and can be computed efficiently.

The above n_2 coefficients are then projected onto another dictionary to produce a new patch. The overlapping patches are then averaged. As discussed above, this is equivalent to linear convolutions on the second feature maps. See the right part of Figure 2.

The above discussion shows that the traditional method can be viewed as a kind of convolutional neural network (with a different non-linear mapping). But not all operations have been considered in the optimization in the traditional methods. On the contrary, in our convolutional neural network, the two dictionaries, non-linear mapping and averaging, are all involved in the filters to be optimized. So our method optimizes an end-to-end mapping that consists of all operations. This is one of the reasons why the CNN gives superior performance.

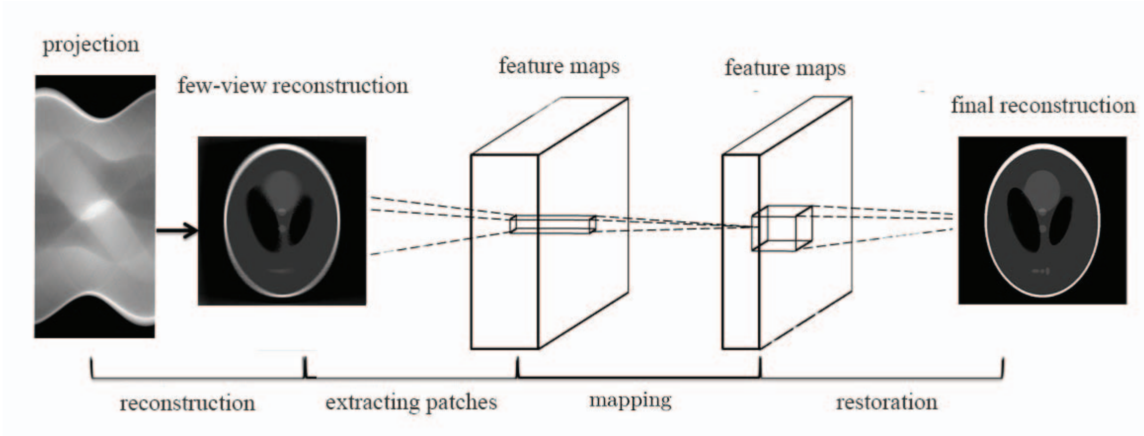

Fig. 1. Given few-view projection data, we can reconstuct few-view images. The first convolutional layer of the CNN extracts a set of feature maps. The second layer maps these feature maps nonlinearly to patch representations with higher image quality. The last layer combines the predictions within a spatial neighbourhood to produce the final reconstruction image.

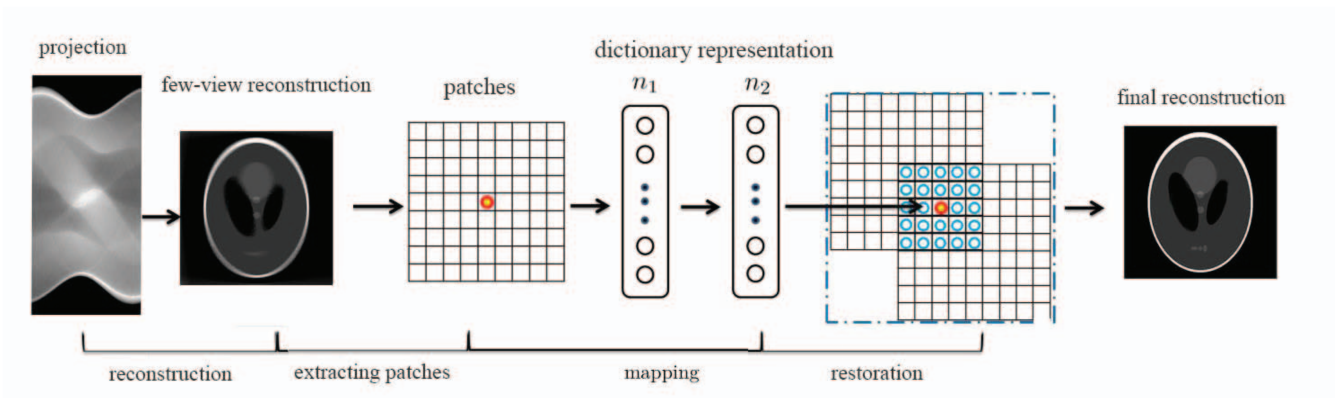

Fig. 2. An illustration of dictionary-learning-based methods in the view of a convolutional neural network.

C. Training

The above discussion shows the convolution neural network is determined by parameters $\Theta = W_1, W_2, W_3, B_1, B_2, B_3$. So training the end-to-end mapping network F requires minimizing the loss between the reconstructed images $F(\mathbf{Y};\Theta)$ and the corresponding ground truth $\mathbf{Y}$.

Given a set of training image pairs $\{\mathbf{X}_i\}$ and $\{\mathbf{Y}_i\}$, we use Mean Squared Error (MSE) as the loss function:

$$\mathcal{L}(\Theta) = \frac{1}{n}\sum_{i=1}^{n} ||F(\mathbf{Y}_i;\Theta) - \mathbf{X}_i||^2 \tag{4}$$

where n is the number of training samples. It is worth noticing that the convolutional neural networks do not preclude the usage of other kinds of loss functions, if only the loss functions are derivable. If a better perceptually motivated metric is given during training, it is flexible for the network to adapt to that metric. On the contrary, such a flexibility is in general difficult to achieve for traditional hand-crafted methods.

In the training phase, the filter weights of each layer are initialized by drawing randomly from a Gaussian distribution with mean 0 and standard deviation 0.001 (and 0 for biases). The learning rate is 10^{-4} . To avoid border effects during training, all the convolutional layers have no padding The MSE loss function is evaluated only by the difference between the central pixels of X_i and the network output.

III. Experimental Results

The experimentation is done on a simulation CT scan. The system matrix in simulation is generated by distance driven algorithm [16]. The geometry is shown as Fig 3: The distance between source and axis is 500mm, and between the detector and the source is 750mm. The detector pixel width is 0.1mm. And the projection data is acquired in 9 angles ranged in 190 degree. We implement our model based on Keras python package using Theano backend. The network is trained on a Titan X GPU. The reconstuction results are shown in Fig 4

From the results III , we can conclude that the proposed method gives an obvious improvement of the image quality, especially for maintaining tiny features(for example,the 3 ellipses below in the image). But the problem is so ill-posed, so the proposed method performs not very good in noisy situation, which needs more work on it.

IV. Conclusions

We propose a deep learning method for few-view CT reconstuction. Our method directly learns an end-to-end mapping

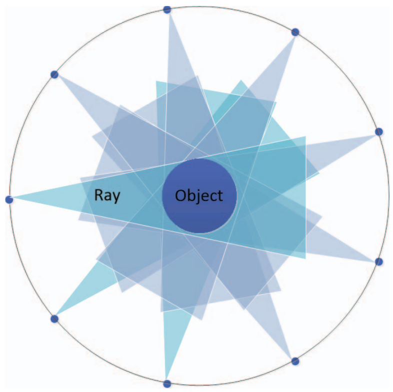

Fig. 3. Scanning geometry of few-view CT with 9 views.

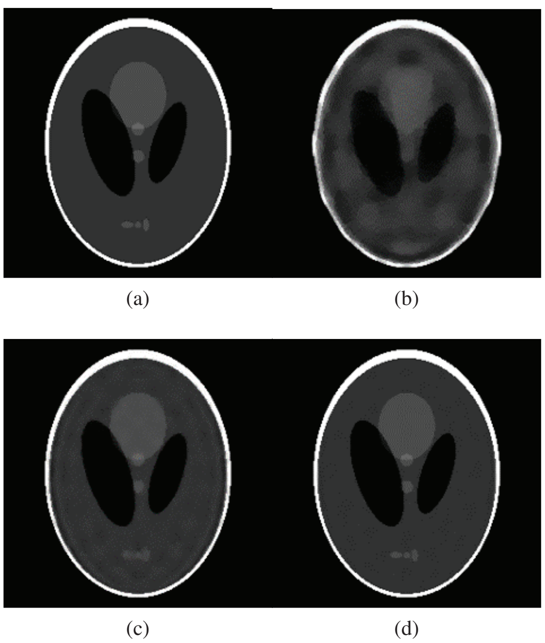

Fig. 4. (a)original phantom (b)ART reconstruction result(5000 iterations) (c) result processed by the proposed network (d) use the proposed network in each ART iterative step

between the few-view/full-view reconstruction. The mapping is represented as a deep convolutional neural network (CNN) that takes the few-view reconstruction image as the input and outputs the full-view one. We further show that traditional Dictionary Learning based reconstruction methods can also be viewed as a deep convolutional network. But unlike traditional methods that handle each component separately, our method jointly optimizes all layers. Our deep CNN has a lightweight structure, yet demonstrates state-of-the-art reconstruction quality, and achieves fast speed for practical on-line usage. We explore different network structures and parameter settings to achieve trade-offs between performance and speed.

Overall, the contributions of this study are mainly in two aspects:

1) We present a fully convolutional neural network for CT few-view reconstruction. The network directly learns an end-to-end mapping between few- and full-view images optimization. The method can be used in both analytic and iterative methods.
2) We establish a relationship between our deep-learning-based CT reconstruction method and the traditional dictionary-learning-based reconstruction methods. This relationship provides a guidance for the design of the network.

REFERENCES

[1] D. J. Brenner and E. J. Hall, "Computed tomographyan increasing source of radiation exposure," *New England Journal of Medicine*, vol. 357, no. 22, pp. 2277–2284, 2007.

[2] D. J. Brenner, C. D. Elliston, E. J. Hall, and W. E. Berdon, "Estimated risks of radiation-induced fatal cancer from pediatric ct," *American journal of roentgenology*, vol. 176, no. 2, pp. 289–296, 2001.

[3] A. B. de Gonzalez and S. Darby, "Risk of cancer from diagnostic x-rays: estimates for the uk and 14 other countries," *The lancet*, vol. 363, no. 9406, pp. 345–351, 2004.

[4] P. J. La Riviere, "Penalized-likelihood sinogram smoothing for low-dose ct," *Medical physics*, vol. 32, no. 6, pp. 1676–1683, 2005.

[5] J. Wang, T. Li, H. Lu, and Z. Liang, "Penalized weighted least-squares approach to sinogram noise reduction and image reconstruction for low-dose x-ray computed tomography," *IEEE transactions on medical imaging*, vol. 25, no. 10, pp. 1272–1283, 2006.

[6] C. Kamphuis and F. J. Beekman, "Accelerated iterative transmission ct reconstruction using an ordered subsets convex algorithm," *IEEE Transactions on Medical Imaging*, vol. 17, no. 6, pp. 1101–1105, 1998.

[7] D. L. Donoho, "Compressed sensing," *IEEE Transactions on information theory*, vol. 52, no. 4, pp. 1289–1306, 2006.

[8] E. J. Candès, J. Romberg, and T. Tao, "Robust uncertainty principles: Exact signal reconstruction from highly incomplete frequency information," *Information Theory, IEEE Transactions on*, vol. 52, no. 2, pp. 489–509, 2006.

[9] J. Song, Q. H. Liu, G. A. Johnson, and C. T. Badea, "Sparseness prior based iterative image reconstruction for retrospectively gated cardiac micro-ct," *Medical physics*, vol. 34, no. 11, pp. 4476–4483, 2007.

[10] H. Yu and G. Wang, "Compressed sensing based interior tomography," *Physics in medicine and biology*, vol. 54, no. 9, p. 2791, 2009.

[11] Z. Tian, X. Jia, K. Yuan, T. Pan, and S. B. Jiang, "Low-dose ct reconstruction via edge-preserving total variation regularization," *Physics in medicine and biology*, vol. 56, no. 18, p. 5949, 2011.

[12] Z. Chen, X. Jin, L. Li, and G. Wang, "A limited-angle ct reconstruction method based on anisotropic tv minimization," *Physics in medicine and biology*, vol. 58, no. 7, p. 2119, 2013.

[13] X. Mou, J. Wu, T. Bai, Q. Xu, H. Yu, and G. Wang, "Dictionary learning based low-dose x-ray ct reconstruction using a balancing principle," in *SPIE Optical Engineering+ Applications*. International Society for Optics and Photonics, 2014, pp. 921 207–921 207.

[14] Q. Xu, H. Yu, X. Mou, and G. Wang, "Dictionary learning based low-dose x-ray ct reconstruction," *Frontiers of Medical Imaging*, p. 99, 2014.

[15] H. Lee, A. Battle, R. Raina, and A. Y. Ng, "Efficient sparse coding algorithms," *Advances in neural information processing systems*, vol. 19, p. 801, 2007.

[16] B. De Man and S. Basu, "Distance-driven projection and backprojection in three dimensions," *Physics in medicine and biology*, vol. 49, no. 11, p. 2463, 2004.