High-resolution CT Image Retrieval Using Sparse Convolutional Neural Network

Yang Lei, Emory University
Dong Xu, Zhejiang Cancer Hospital
Zhengyang Zhou, Emory University
Kristin Higgins, Emory University
Xueyuan Dong, Emory University
Tian Liu, Emory University
Hyunsuk Shim, Emory University
Hui Mao, Emory University
Walter J Curran, Emory University
Xiaofeng Yang, Emory University

Journal Title: Proceedings of SPIE
Volume: Volume 10573
Publisher: Society of Photo-optical Instrumentation Engineers (SPIE) | 2018-03-09
Type of Work: Article | Final Publisher PDF
Publisher DOI: 10.1117/12.2292891
Permanent URL: https://pid.emory.edu/ark:/25593/v79z4

Final published version: http://dx.doi.org/10.1117/12.2292891

Copyright information:
© (2018) COPYRIGHT Society of Photo-Optical Instrumentation Engineers (SPIE). Downloading of the abstract is permitted for personal use only.

Accessed April 24, 2020 11:47 PM EDT
High-resolution CT image retrieval using sparse convolutional neural network

Lei, Yang, Xu, Dong, Zhou, Zhengyang, Higgins, Kristin, Dong, Xue, et al.
High-resolution CT Image Retrieval Using Sparse Convolutional Neural Network

Yang Leia, Dong Xub, Zhengyang Zhouc, Kristin Higginsa, Xue Donga, Tian Liua, Hyunsuk Shimad,
Hui Moad, Walter J. Currania and Xiaofeng Yanga, *

a Department of Radiation Oncology and Winship Cancer Institute, Emory University, Atlanta, GA 30322
b Department of Ultrasound Imaging, Zhejiang Cancer Hospital, Hangzhou, China 310022
c Department of Radiology, Nanjing Drum Tower Hospital, the Affiliated Hospital of Nanjing University Medical School, Nanjing, China 210008
d Department of Radiation Oncology and Imaging Sciences and Winship Cancer Institute, Emory University, Atlanta, GA 30322

*Corresponding author: xiaofeng.yang@emory.edu

ABSTRACT

We propose a high-resolution CT image retrieval method based on sparse convolutional neural network. The proposed framework is used to train the end-to-end mapping from low-resolution to high-resolution images. The patch-wise feature of low-resolution CT is extracted and sparsely represented by a convolutional layer and a learned iterative shrinkage threshold framework, respectively. Restricted linear unit is utilized to non-linearly map the low-resolution sparse coefficients to the high-resolution ones. An adaptive high-resolution dictionary is applied to construct the informative signature which is highly connected to a high-resolution patch. Finally, we feed the signature to a convolutional layer to reconstruct the predicted high-resolution patches and average these overlapping patches to generate high-resolution CT. The loss function between reconstructed images and the corresponding ground truth high-resolution images is applied to optimize the parameters of end-to-end neural network. The well-trained map is used to generate the high-resolution CT from a new low-resolution input. This technique was tested with brain and lung CT images and the image quality was assessed using the corresponding CT images. Peak signal-to-noise ratio (PSNR), structural similarity index (SSIM) and mean absolute error (MAE) indexes were used to quantify the differences between the generated high-resolution and corresponding ground truth CT images. The experimental results showed the proposed method could enhance images resolution from low-resolution images. The proposed method has great potential in improving radiation dose calculation and delivery accuracy and decreasing CT radiation exposure of patients.

Keywords: High-resolution image retrieval, convolutional neural network, CT.

1. INTRODUCTION

Due to CT x-ray exposure and respiratory motion, a routine planning CT (lung or abdomen) is usually captured with a large slice thickness (e.g. 3-4mm) (1-5). Since the out-of-slice resolution is distinctly lower than the in-slice resolution in these CT images, such CT images will affect both contouring (tumor and organ) and dose calculation in treatment planning (6-10). Thus, CT images with super-resolution or high-resolution are needed for the radiotherapy (3, 9, 11-16). The purpose of this work is to develop a deep-learning-based method to retrieve/reconstruct high-resolution CT images from routine low-resolution CT for radiotherapy treatment planning.
2. METHOD

Suppose we have a set of pairs of low-resolution and high-resolution training CT images. For each pair, the high-resolution image is used as the mapping target of the low-resolution image. The mapping is represented as a deep sparse convolutional neural network (17-20) that takes the low-resolution image as the input and outputs the high-resolution one. In the training stage, we first upscale low-resolution to the high-resolution size by the bicubic interpolation, then extract features of up-scaled image. A classical feature extraction strategy in image restoration is to densely extract patches and then represent them by a set of pre-trained bases such as PCA (21), DCT (22), Haar (23), etc. This is equivalent to convolving the image by a set of filters, each of which is a basis. Thus, the input up-scaled low-resolution image $I_{\text{LR}}$ first goes through a convolutional layer $H_i(I_{\text{LR}}) = \max(0, F_i \ast I_{\text{LR}} + B_i)$ to extract features for each patch, where $F_i$ denotes the filters and $B_i$ denotes the biases, $\ast$ represents the convolution operation. Secondly, the learned iterative shrinkage threshold algorithm (LISTA) (24-26) layer $H_i$ is applied to adaptively sparse represent the extracted feature by iteratively optimize the low-resolution dictionary $D_{\text{LR}}$ and high-dimensional sparse code $h_i(H_i(I_{\text{LR}}))$. Thirdly, in order to construct the accurate mapping from low-resolution to high-resolution, the obtained sparse vector is mapped to the sparse coefficient of high-resolution by a non-linear mapping based on restricted linear unit (ReLU) method (27). Then, we use a high-resolution dictionary $D_{\text{HR}}$ to reconstruct the dense features of high-resolution from sparse coefficients and then feed the features into the final layer $H_4$ to reconstruct the high-resolution patches. To obtain the adaptive $D_{\text{HR}}$, we first randomly set $D_{\text{HR}}$ with Gaussian noise, and then optimize it by iterative shrinkage threshold algorithm (ISTA) (28, 29). The predicted overlap patches are averaged to produce the final full image, which can be regarded as a pre-defined filter on a set of feature maps (where each position is the “flattened” vector form of a high-resolution patch). Motivated by this, we define a convolutional layer to produce the final high-resolution image $H_4(\mathcal{P}(I_{\text{LR}})) = \max(0, F_4 \ast \mathcal{P}(I_{\text{LR}}) + B_4)$, where $\mathcal{P}$ denotes the previous operations, $F_4$ is composed of several filters and $B_4$ is a bias vector. The framework of the proposed algorithm is shown in Fig. 1.

![Figure 1. Schematic flow chart of the proposed algorithm for high-resolution CT image restoration.](image-url)
2.1 Learned Iterative Shrinkage Threshold Algorithm

Our proposed sparse operator $H_2$ is based on the intimate connection between sparse coding and neural network studied by (24). After feature extraction, for a given input vector $x = H_1(l_{in}) \in \mathbb{R}^n$ with fixed low-resolution dictionary $D_{lx}$, the goal is to find the optimal sparse code vector $z \in \mathbb{R}^n$ which minimizes an energy function that combines the square reconstruction error and a $l_1$ sparsity penalty on the code:

$$\arg \min_z \frac{1}{2} \| D_{lx} z \|^2 + \alpha \| z \|_1$$

(1)

where $D_{lx}$ is an $n \times m$ dictionary matrix whose columns are the normalized basis vectors, $\alpha$ is a coefficient controlling the sparsity penalty. In order to adaptively optimize the sparse coefficient, we first initiate the low-resolution dictionary by randomly set $D_{lx}(0)$ with Gaussian noise. Then we iteratively optimize the dictionary and sparse code by LISTA:

$$z_{k+1} = h_b \left( \frac{1}{L} D_{lx} x + (1 - \frac{1}{L} D_{lx}^\top(k) D_{lx}(k)) z_k \right), D_{lx}(k + 1) = \arg \min_{D_{lx}} \frac{1}{2} \| z - D_{lx}(k) \|^2 + \alpha \| z \|_1$$

(2)

where $h_b$ denotes the shrinkage function $[h_b(v)] \rightarrow \text{sign}(v) |v| - \theta$, $\theta = \alpha / L$, $L$ is the upper bound on the largest eigenvalue of $D_{lx}^\top(k) D_{lx}(k)$.

2.2 Restricted Linear Unit

Previous layers extract a $n_i$-dimensional sparse representation for each low-resolution patch. In order to enhance the performance of neural network, it is demanded to learn the connection of low-resolution sparse representations with high-resolution sparse codes. In other word, it is needed to obtain the informative and relevant representation which can be highly connected to the patch of high-resolution from previous low-resolution representation. We perform a non-linear mapping to build the $n_2$-dimensional high-resolution vectors from $n_i$-dimensional. This is equivalent to applying filters which have a trivial spatial support 1×1:

$$H_2(z) = \max(0, F_2 * x + B_2)$$

(3)

Here $F_2$ contains $n_2$ filters of size $n_i \times 1 \times 1$, $B_2$ is $n_2$-dimensional. Each of the output $n_2$-dimensional vectors are conceptually a sparse representation of a high-resolution patch that will be used for reconstruction. Rectified Linear Unit (ReLU, $\max(0, x)$) (27) is used on the filter responses. It is possible to add more convolutional layers to increase the non-linearity. But this can increase the complexity of the model ($n_2 \times 1 \times 1 \times n$ parameters for one layer), and thus demands more training time.

2.3 Loss Function

Learning the end-to-end mapping function $F_{\text{CNN}}$ requires the estimation of all the network parameters $W = \{F_1, B_1, F_2, B_2, D_{lx}, \alpha, F_3, B_3, \cdots \}$. This is achieved through minimizing the loss between the reconstructed images $F_{\text{CNN}}(l_{in}, W)$ and the corresponding ground truth high-resolution images $l_{in}$. Given a set of high-resolution images $\{l_{in}^i\}$ and their corresponding low-resolution images $\{l_{lx}^i\}$, we use mean squared error (MSE) as the loss function:

$$L(W) = \frac{1}{n} \sum_{i=1}^{n} \| F_{\text{CNN}}(l_{lx}^i, W) - l_{in}^i \|^2$$

(4)

where $n$ is the number of training samples. The parameters of each convolutional layer are initialized by drawing randomly from a Gaussian distribution with zero mean and standard deviation 0.001 and 0 for biases. The MSE of loss function is evaluated by the difference between the $\{l_{in}^i\}$ and the network output. We use the Caffe package (30) to enhance the training performance.
3. RESULTS
The proposed high-resolution CT retrieve method was tested with a dataset with 50 brain and lung CT images with the pixel size was 1.00×1.00×2.00 mm$^3$ and 0.98×0.98×3.00 mm$^3$. We performed leave-one-out cross-validation method to evaluate the proposed high-resolution reconstruction algorithm. We repeated the training and testing for our method twice using two different resolution CT images down-sampled by 2 and 3 from original brain and lung CT images. Our retrieved CT images (output) were compared with the original planning CT images. In order to get a quantitative evaluation, we used the mean absolute error (MAE), peak signal-to-noise ratio (PSNR) and structural similarity index (SSIM) to evaluate the difference between original CT (gold standard) and our reconstructed CT.

Fig. 2 shows a comparison of CT images retrieved from different low-resolution CT images (down-sampled by 2 and 3) using Bicubic interpolation, Dong et al's method (31) and our proposed method. The retrieved CT images by our method are much closer to original CT images. Table 1 and Table 2 shows average MAE, PSNR and SSIM of three methods for all brain and lung patient's data. The smaller MAE, as well as higher PSNR and SSIM demonstrated the restoration accuracy of the proposed method.
Table 2. MAE, PSNR and SSIM for the brain CT images using the three methods.

<table>
<thead>
<tr>
<th>Method</th>
<th>MAE</th>
<th>PSNR (dB)</th>
<th>SSIM</th>
<th>MAE</th>
<th>PSNR (dB)</th>
<th>SSIM</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bicubic</td>
<td>19.75</td>
<td>33.04</td>
<td>0.97</td>
<td>35.11</td>
<td>28.62</td>
<td>0.91</td>
</tr>
<tr>
<td>Dong et al. (31)</td>
<td>14.90</td>
<td>35.74</td>
<td>0.98</td>
<td>27.10</td>
<td>30.77</td>
<td>0.94</td>
</tr>
<tr>
<td>The proposed</td>
<td>9.51</td>
<td>38.71</td>
<td>0.99</td>
<td>20.08</td>
<td>32.32</td>
<td>0.96</td>
</tr>
</tbody>
</table>

4. DISCUSSION AND CONCLUSION

In this paper, we propose a high-resolution CT retrieve method based on sparse convolutional neural network. The novelty of our approach is the integration of the sparse representation and deep convolution neural network into the high-resolution CT reconstruction framework. This approach has 2 distinctive strengths: 1) Instead of using conventional sparse representation, learnable network-based dictionary learning and sparse representation is used to sparse represent the low-resolution image. 2) Contrary to the classical deep networks for image super-resolution with sparse prior, which uses sparse coefficients as the representation of both low-resolution and high-resolution, we used a ReLU-based non-linear mapping to build the informative and relevant representation that is highly connected to the patch of high-resolution. The parameters of each layer include the non-linear mapping are optimized by loss function iteratively. Thus, we build an adaptive connection between low- and high-resolution CT images. We compared the proposed method with the state-art-of methods and demonstrated its feasibility and reliability. The proposed method has great potential in improving radiation dose calculation and delivery accuracy, and decreasing CT radiation exposure of patients.

ACKNOWLEDGMENT

This research is supported in part by the National Cancer Institute of the National Institutes of Health under Award Number R01CA215718, the Department of Defense (DoD) Prostate Cancer Research Program (PCRP) Award W81XWH-13-1-0269 and Dunwoody Golf Club Prostate Cancer Research Award, a philanthropic award provided by the Winship Cancer Institute of Emory University.

REFERENCES

