Published in final edited form as:

Proc SPIE Int Soc Opt Eng. 2018 February ; 10573: . doi:10.1117/12.2292891.

High-resolution CT Image Retrieval Using Sparse Convolutional Neural Network

Yang Lei^a, Dong Xu^b, Zhengyang Zhou^c, Kristin Higgins^a, Xue Dong^a, Tian Liu^a, Hyunsuk Shim^{a,d}, Hui Mao^d, Walter J. Curran^a, Xiaofeng Yang^{a,*}

^aDepartment of Radiation Oncology and Winship Cancer Institute, Emory University, Atlanta, GA 30322

^bDepartment of Ultrasound Imaging, Zhejiang Cancer Hospital, Hangzhou, China 310022

^cDepartment of Radiology, Nanjing Drum Tower Hospital, the Affiliated Hospital of Nanjing University Medical School, Nanjing, China 210008

^dDepartment of Radiation Oncology and Imaging Sciences and Winship Cancer Institute, Emory University, Atlanta, GA 30322

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

*Corresponding author: xiaofeng.yang@emory.edu.

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_{LR} first goes through a convolutional layer $H_1(I_{LR}) = \max(0, F_1 * I_{LR} + B_1)$ to extract features for each patch, where F_1 denotes the filters and B_1 denotes the biases, $*$ represents the convolution operation. Secondly, the learned iterative shrinkage threshold algorithm (LISTA) (24-26) layer H_2 is applied to adaptively sparse represent the extracted feature by iteratively optimize the low-resolution dictionary D_{LR} and high-dimensional sparse code $H_2(H_1(I_{LR}))$. 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 H_3 based on restricted linear unit (ReLU) method (27). Then, we use a high-resolution dictionary D_{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_{HR} , we first randomly set D_{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 $H_4(\tilde{H}(I_{LR})) = \max(0, F_4 * \tilde{H}(I_{LR}) + B_4)$, where $\tilde{H}$ 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.

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 $x = H_1(I_{LR}) \in \mathbb{R}^n$ with fixed low-resolution dictionary D_{LR} , the goal is to find the optimal

sparse code vector $z \in R^m$ which minimizes an energy function that combines the square reconstruction error and a l_1 sparsity penalty on the code:

$$\operatorname{argmin}_z \frac{1}{2} \|x - D_{LR}z\|_2^2 + \alpha \|z\|_1 \quad (1)$$

where D_{LR} is an $n \times m$ dictionary matrix whose columns are the normalized basis vectors, α 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 $D_{LR}(0)$ with Gaussian noise. Then we iteratively optimize the dictionary and sparse code by LISTA:

$$z_{k+1} = h_{\theta} \left(\frac{1}{L} D_{LR}(k)x + \left(1 - \frac{1}{L} D_{LR}^T(k) D_{LR}(k)\right) z_k \right), D_{LR}(k+1) = \operatorname{argmin}_{D_{LR}} \frac{1}{2} \|x - D_{LR}(k)z\|_2^2 + \alpha \|z_{k+1}\|_1 \quad (2)$$

where h_{θ} denotes the shrinkage function $[h_{\theta}(v)]_i = \operatorname{sign}(v_i)(|v_i| - \theta_i)$, $\theta_i = \alpha / L$, L is the upper bound on the largest eigenvalue Of $D_{LR}^T(k) D_{LR}(k)$.

2.2 Restricted Linear Unit

Previous layers extract a n_1 – 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_1 – dimensional. This is equivalent to applying filters which have a trivial spatial support 1×1 :

$$H_3(z) = \max(0, F_3 * x + B_3) \quad (3)$$

Here F_2 contains n_2 filters of size $n_1 \times 1 \times 1$, 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 ($n_2 \times 1 \times 1 \times n_2$ parameters for one layer), and thus demands more training time.

2.3 Loss Function

Learning the end-to-end mapping function F_{SCNN} requires the estimation of all the network parameters $W = \{F_1, B_1, D_{LR}, \alpha, F_3, B_3, D_{HR}, F_4, B_4\}$. This is achieved through minimizing the loss between the reconstructed images $F_{\text{SCNN}}(I_{LR}, W)$ and the corresponding ground truth

high-resolution images I_{HR} . Given a set of high-resolution images $\{I_{HR}^i\}$ and their corresponding low-resolution images $\{I_{LR}^i\}$, we use mean squared error (MSE) as the loss function:

$$L(W) = \frac{1}{n} \sum_{i=1}^n \|F_{SCNN}(I_{LR}^i, W) - I_{HR}^i\|_2^2 \quad (4)$$

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 $\{I_{HR}^i\}$ 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 $1.00 \times 1.00 \times 2.00 \text{ mm}^3$ and $0.98 \times 0.98 \times 3.00 \text{ mm}^3$. 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.

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

1. Yang X, Liu T, Dong X, Tang X, Elder E, Curran WJ, and Dhabaan A, "A patch-based CBCT scatter artifact correction using prior CT," *Proc. SPIE* 10132(1013229 (2017).
2. Yang X, Lei Y, Shu H-K, Rossi P, Mao H, Shim H, Curran WJ, and Liu T, "Pseudo CT estimation from MRI using patch-based random forest," *Proc. SPIE* 10133(101332Q-101337 (2017).
3. Yang XF, Wu SY, Sechopoulos I, and Fei BW, "Cupping artifact correction and automated classification for high-resolution dedicated breast CT images Xiaofeng Yang and Shengyong Wu," *Medical Physics* 39(10), 6397–6406 (2012). [PubMed: 23039675]
4. Yang X, Wang T, Liu T, Zhu L, Khan MK, El-rayes B, Curran WJ, and Landry JC, "Optimal Treatment Phase Selection Based on a 4DCT Deformable Registration for Stereotactic Body Radiation Therapy in Pancreatic Cancer," *International Journal of Radiation Oncology*Biophysics* 99(2), E743 (2017).
5. Yang X, Lei Y, Shu HKG, Rossi PJ, Mao H, Shim H, Curran WJ Jr., and Liu T, "A Learning-Based Approach to Derive Electron Density from Anatomical MRI for Radiation Therapy Treatment Planning," *International Journal of Radiation Oncology · Biology · Physics* 99(2), S173–S174 (2017).
6. Kazerooni EA, "High-Resolution CT of the Lungs," *American Journal of Roentgenology* 177(3), 501–519 (2001). [PubMed: 11517038]
7. Mayo JR, Jackson SA, and Muller NL, "High-resolution CT of the chest: radiation dose," *AJR Am J Roentgenol* 160(3), 479–481 (1993). [PubMed: 8430539]
8. Yang XF, Wu N, Cheng GH, Zhou ZY, Yu DS, Beitler JJ, Curran WJ, and Liu T, "Automated Segmentation of the Parotid Gland Based on Atlas Registration and Machine Learning: A Longitudinal MRI Study in Head-and-Neck Radiation Therapy," *Int J Radiat Oncol* 90(5), 1225–1233 (2014).
9. Yang X, Rossi P, Ogunleye T, Marcus DM, Jani AB, Mao H, Curran WJ, and Liu T, "Prostate CT segmentation method based on nonrigid registration in ultrasound-guided CT-based HDR prostate brachytherapy," *Med Phys* 41(11), 111915 (2014). [PubMed: 25370648]
10. Yang XF, Tridandapani S, Beitler JJ, Yu DS, Yoshida EJ, Curran WJ, and Liu T, "Ultrasound GLCM texture analysis of radiation-induced parotid-gland injury in head-and-neck cancer radiotherapy: An in vivo study of late toxicity," *Medical Physics* 39(9), 5732–5739 (2012). [PubMed: 22957638]
11. Yan Z, Li J, Lu Y, Yan H, and Zhao Y, "Super resolution in CT," *International Journal of Imaging Systems and Technology* 25(1), 92–101 (2015).
12. Hakimi WE, and Wesarg S, "Accurate super-resolution reconstruction for CT and MR images," *Proceedings of the 26th IEEE International Symposium on Computer-Based Medical Systems* 445–448 (2013).
13. Aarle W. v., Batenburg KJ, Gompel GV, Castelee E. V. d., and Sijbers J, "Super-Resolution for Computed Tomography Based on Discrete Tomography," *IEEE Transactions on Image Processing* 23(3), 1181–1193 (2014). [PubMed: 24723522]

14. Shiting F, Huafeng W, Yueliang L, Minghui Z, Wei Y, Qianjin F, Wufan C, and Yu Z, "Super-resolution reconstruction of 4D-CT lung data via patch-based low-rank matrix reconstruction," *Physics in Medicine & Biology* 62(20), 7925 (2017). [PubMed: 28872050]
15. Yang XF, and Fei BW, "A multiscale and multiblock fuzzy C-means classification method for brain MR images," *Medical Physics* 38(6), 2879–2891 (2011). [PubMed: 21815363]
16. Yang X, Ghafourian P, Sharma P, Salman K, Martin D, and Fei B, "Nonrigid Registration and Classification of the Kidneys in 3D Dynamic Contrast Enhanced (DCE) MR Images," *Proc SPIE Int Soc Opt Eng* 8314(83140B) (2012).
17. Dong C, Loy CC, He KM, and Tang XO, "Image Super-Resolution Using Deep Convolutional Networks," *Ieee Transactions on Pattern Analysis and Machine Intelligence* 38(2), 295–307 (2016). [PubMed: 26761735]
18. Baoyuan L, Min W, Foroosh H, Tappen M, and Pensky M, "Sparse Convolutional Neural Networks," 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR) 806–814 (2015).
19. Parashar A, Rhu M, Mukkara A, Puglielli A, Venkatesan R, Khailany B, Emer J, Keckler SW, and Dally WJ, "SCNN: An Accelerator for Compressed-sparse Convolutional Neural Networks," *SIGARCH Comput. Archit. News* 45(2), 27–40 (2017).
20. Graham B, Spatially-sparse convolutional neural networks (2014).
21. Rejichi S, and Chaabane F, "Feature Extraction Using Pca for Vhr Satellite Image Time Series Spatio-Temporal Classification," 2015 Ieee International Geoscience and Remote Sensing Symposium (Igarss) 485–488 (2015).
22. Shobana G, and Balakrishnan R, "Brain tumor diagnosis from MRI feature analysis - A comparative study," 2015 International Conference on Innovations in Information, Embedded and Communication Systems (ICIIECS) 1–4 (2015).
23. Smitha JC, and Babu SS, "MRI Brain Image Classification Using Haar Wavelet and Artificial Neural Network," in *Artificial Intelligence and Evolutionary Algorithms in Engineering Systems: Proceedings of ICAEES 2014*, Volume 2 Suresh LP, Dash SS, and Panigrahi BK, Eds., pp. 253–261, Springer India, New Delhi (2015).
24. Gregor K, and LeCun Y, "Learning fast approximations of sparse coding," in *Proceedings of the 27th International Conference on Machine Learning*, pp. 399–406, Omnipress, Haifa, Israel (2010).
25. Beck A, and Teboulle M, "A Fast Iterative Shrinkage-Thresholding Algorithm for Linear Inverse Problems," *SIAM Journal on Imaging Sciences* 2(1), 183–202 (2009).
26. Zhang J, and Ghanem B, "ISTA-Net: Iterative Shrinkage-Thresholding Algorithm Inspired Deep Network for Image Compressive Sensing," *CoRR* abs/1706.07929((2017).
27. Nair V, and Hinton GE, "Rectified Linear Units Improve Restricted Boltzmann Machines," pp. 807–814 (2010).
28. Daubechies I, Defrise M, and De Mol C, "An iterative thresholding algorithm for linear inverse problems with a sparsity constraint," *Communications on Pure and Applied Mathematics* 57(11), 1413–1457 (2004).
29. Mourad N, and Reilly JP, "Automatic Threshold Estimation for Iterative Shrinkage Algorithms Used with Compressed Sensing," 2012 Ieee International Conference on Acoustics, Speech and Signal Processing (Icassp) 2721–2724 (2012).
30. Jia Y, Shelhamer E, Donahue J, Karayev S, Long J, Girshick R, Guadarrama S, and Darrell T, "Caffe: Convolutional Architecture for Fast Feature Embedding," in *Proceedings of the 22nd ACM international conference on Multimedia*, pp. 675–678, ACM, Orlando, Florida, USA (2014).
31. Dong W, Zhang L, Shi G, and Wu X, "Image Deblurring and Super-Resolution by Adaptive Sparse Domain Selection and Adaptive Regularization," *IEEE Transactions on Image Processing* 20(7), 1838–1857 (2011). [PubMed: 21278019]

Figure 1.
Schematic flow chart of the proposed algorithm for high-resolution CT image restoration.

Figure 2.

Comparison of high-resolution CT images retrieved from different low-resolution CT images using three different methods. From left to right (a) is the ground truth CT, (b) is the high-resolution CT obtained by bicubic interpolation, (c) is the difference between the bicubic-based high-resolution and ground truth CT, (d) is the high-resolution CT obtained from a published state-of-art method [11], (e) is the difference between the high-resolution CT obtained by this method and ground truth CT, (f) is the high-resolution CT obtained by the proposed method, (g) is the difference between the high-resolution CT obtained by our proposed method and ground truth CT. The first and third rows show the results of CT images down-sampled by 2. The second and fourth rows show the results of CT images down-sampled by 3.

Table 1.

MAE, PSNR and SSIM for the lung CT images using the three methods.

	Down-sampling factor = 2			Down-sampling factor = 3		
	MAE	PSNR (dB)	SSIM	MAE	PSNR (dB)	SSIM
Bicubic	18.61	34.49	0.94	28.05	30.91	0.89
Dongei <i>et al.</i> (31)	15.42	36.71	0.96	22.82	33.03	0.91
The proposed	13.64	37.93	0.96	21.16	34.40	0.92

Table 2.

MAE, PSNR and SSIM for the brain CT images using the three methods.

	Down-sampling factor = 2			Down-sampling factor = 3		
	MAE	PSNR (dB)	SSIM	MAE	PSNR (dB)	SSIM
Bicubic	19.75	33.04	0.97	35.11	28.62	0.91
Dongei <i>et al.</i> (31)	14.90	35.74	0.98	27.10	30.77	0.94
The proposed	9.51	38.71	0.99	20.08	32.32	0.96