COMPUTED TOMOGRAPHY SUPER-RESOLUTION USING CONVOLUTIONAL NEURAL NETWORKS

Haichao Yu[*][†], *Ding Liu*[*][†], *Honghui Shi*[*][†], *Hanchao Yu*[†], *Zhangyang Wang*[‡],
Xinchao Wang[†], *Brent Cross*[§], *Matthew Bramlet*[♭], *Thomas S. Huang*[†]

†Beckman Institute, University of Illinois at Urbana-Champaign
‡Department of Computer Science and Engineering, Texas A&M University
§Jump Trading Simulation and Education Center ♭ University of Illinois College of Medicine

ABSTRACT

The practical application of Computed Tomography (CT) faces the dilemma between higher image resolution and less X-ray exposure for patients, motivating the research on CT super-resolution (SR). In this paper, we apply state-of-the-art SR techniques to reconstruct CT images using two proposed advanced CT SR models based on Convolutional Neural Networks (CNNs) and residual learning: a single-slice CT SR network (S-CTSRN), and a multi-slice CT SR network (M-CTSRN). S-CTSRN improves the high-frequency feature extraction by incorporating the residual learning strategy, while M-CTSRN further utilizes the coherence between neighboring CT slices for better SR reconstruction. We evaluate both models on a large-scale CT dataset[1], and obtain competitive results both quantitatively and qualitatively.

Index Terms— Super-resolution (SR), Medical Image Analysis, Computed Tomography (CT), Convolutional Neural Network (CNN), Residual Learning

1. INTRODUCTION

Computed Tomography (CT) has been a critical technique for medical diagnosis and decision making [1]. It relies on the invisible X-rays to image bones and soft tissues of patients, after which the received signals are reconstructed as multi-slice CT image sequences to be used to support clinical decisions of medical practitioners.

CT images of higher resolutions enable more accurate medical abnormality detection and are thus highly desirable. In general, there are three ways to improve the resolution of CT images: improving the imaging sensors, refining the reconstruction algorithms, and enhancing the images after reconstruction [2]. However, to obtain higher-resolution CT images using the first two hardware-related methods either introduces higher complexity and more costs for current imaging systems or implies higher-dose exposure of X-rays for patients that can cause the potential risk of adverse health effects due to the need of acquiring better raw data [2]. For instance, the recent study [3] revealed the increased possibility of cancer induction from X-ray radiation exposure. As such, to obtain high-resolution CT images and to maintain low CT radiation dose make a contradiction, which calls for enhancing the resolution of CT images using image processing techniques such as super-resolution (SR) [4, 5, 6].

In this paper, we focus on applying and improving SR techniques on CT image sequences. We propose two advanced SR models for CT images: the single-slice CT SR network (S-CTSRN), and the multi-slice CT SR network (M-CTSRN). Both models are based on the popular and successful Convolutional Neural Networks (CNNs). The S-CTSRN model adopts the residual learning strategy [7] in order to extract richer high-frequency details, and the M-CTSRN model further accounts for the notable coherence between multiple neighboring CT slices for enhanced SR reconstruction.

Our contribution is, therefore, the introduction of state-of-the-art learning-based SR techniques on a large-scale medical CT image dataset to enhance the image resolution. Based on CNNs, the proposed models emphasize learning the residual high-frequency features, as well as taking into account the sequential coherence between CT slices. Our proposed models are evaluated on a large CT dataset[1], with highly competitive performance achieved both quantitatively and qualitatively.

2. RELATED WORK

SR has a wide domain of applications ranging from computer vision [8, 9] to medical imaging [10]. In this section, we first briefly review generic SR approaches and then elaborate SR methods for medical imaging.

2.1. Generic Super-resolution

SR aims at reconstructing high-resolution (HR) images from low-resolution (LR) ones. Apart from interpolation-based

*denotes equal contribution.

[1] We are considering the possibility of opening the CT dataset to public in the future, please contact the authors if you are interested. We would like to thank the generous support from JUMP ARCHES to make this work possible.

methods that are often used as comparison baselines, SR algorithms can be categorized into traditional model-based ones and learning-based ones. Traditional model-based SR algorithms explicitly model the image downsampling degradation process and regularize the upsampling SR reconstruction with various priors [11]. Learning-based SR algorithms, on the other hand, learn representations from large training databases of HR and LR image pairs [12, 13], exploit self-similarities within an image [14], or combine both ways [15].

In the past few years, the success of Convolution Neural Networks (CNNs) in computer vision tasks [16], especially their hierarchical feature extraction and representation capability, has motivated a flood of CNN-based image SR models [17, 18, 19, 20] and video SR models [21, 22]. CNN-based models benefit from end-to-end learning on large-scale datasets, and thus have been renovating the state-of-the-art SR performance. Moreover, once a CNN-based SR model is trained, conducting SR on an image is purely a feed-forward process, making it as well appealing in terms of efficiency.

More recently, deep residual networks [7] further push forward the performance of many visual recognition tasks such as classification and segmentation by a large margin [7]. By introducing residual connections into conventional CNN frameworks, residual networks effectively improve the optimization result for much deeper networks by preventing gradient vanishing and enabling the information flow across skip-connected layers, and hence yielding the superior performance. It is therefore natural to consider incorporating residual networks into the SR scenario.

2.2. Super-resolution for Medical Imaging

SR for medical imaging has been a well-studied specialized field with respect to different imaging modalities, with Magnetic Resonance Imaging (MRI) and CT being the two most representative examples.

To conduct SR on MRI images, Rueda et al. [23] exploited dictionary-based models to generate an HR image from a single LR one. Shi et al. [5] proposed a multi-atlas patch matching algorithm for MRI SR. Towards the reconstruction of temporal sequences and 3D data, Plenge et al. [24] interpolated between MRI slices via the non-local means algorithm [25], without relying on accurate motion estimation and alignment. Poot et al. [4] and Odille et al. [6] reconstructed HR isotropic 3D MRI data from multiple LR MRI slices of different orientations. Oktay et al. [26] introduced deep neural networks that take advantage of both short-axis and long-axis MRI slices for HR MRI data reconstruction.

Compared to SR for MRI, applying SR to CT data appears to be significantly more challenging. All CT slices are of the same orientation, without multi-view information available. The prior work [27] utilized 4D-CT data to reconstruct HR images, which involves several frames for each CT slice at different respiratory phases. Such a setting was different from the common 3D-CT format and also caused more radiation dose for patients. While most literature on CT SR remains to rely on traditional model-based methods, e.g., [28], there have been limited efforts in exploring the usage of deep learning models for CT SR.

3. PROPOSED MODELS

Deep CNN models have been widely used in generic image SR and video SR [17, 29, 22], achieving state-of-the-art SR performance on natural images. Inspired by [22], we incorporate a residual module to learn the high-frequency details of the image, which are then fused with the upsampled image to reconstruct the final HR image with higher fidelity. In the multi-slice model, we deem the consecutive CT slices similar to the temporal frames of a video, which share inherent correlations and can be jointly exploited for SR.

3.1. S-CTSRN: Single-slice CT SR Network

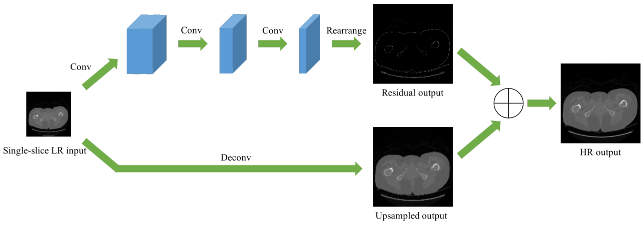


Fig. 1. S-CTSRN, the single-slice CT SR network (zoom in to see details; best viewed in color)

As shown in Fig. 1, the single-slice image-based SR network (S-CTSRN) consists of two branches. The top branch gets inspired by the efficient sub-pixel convolutional neural network (ESPCN) [22]. It is composed of three convolution layers and a special *rearrange* layer. The convolution layers extract useful information from LR input, to be mapped to the HR features. For the output feature map of the last convolutional layer, each channel can be viewed as an LR version of the target HR image, with a certain portion of details reconstructed/enhanced (readers of interests may refer to [22] for details).The *rearrange* layer then flattens and fuses all those feature maps, resulting in one entire residual image containing mostly high-frequency details. The residual image, combined with the base image upsampled from the LR input by a bilinear interpolation layer (the bottom branch in Fig. 1), constitutes the final HR output image.

Reconstructing the high-frequency components is known to be the crucial part in any SR model. By modeling the high-frequency reconstruction using the dedicated residual learning module, the S-CTSRN model shows to be capable to produce sharper edges and finer details.

3.2. M-CTSRN: Multi-slice CT SR Network

Built on S-CTSRN, the multi-slice SR model, M-CTSRN, is also composed of two branches as shown in Fig. 2. To jointly

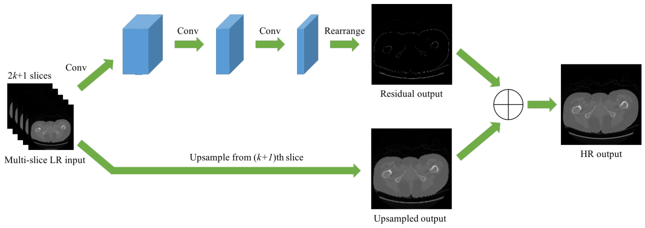


Fig. 2. M-CTSRN, the multi-slice CT SR Network (zoom in to see details; best viewed in color)

utilize consecutive slice information when reconstructing the current frame, we stack multiple adjacent CT slices, centered at the current frame, as input to the top residual branch in M-CTSRN, which treats the consecutive CT slices in the same way as video frames. The rest part of this branch is the same as that of S-CTSRN. For the bottom upsampling branch, we feed only the current frame.

In short, M-CTSRN estimates the residual image using not only the current slice, but also its neighboring slices as side information. The base image remains to be upsampled from the current slice. As a result, M-CTSRN achieves the joint utilization of 3D spatial consistency, producing higher quality SR results.

4. EXPERIMENTS

4.1. The CT Dataset

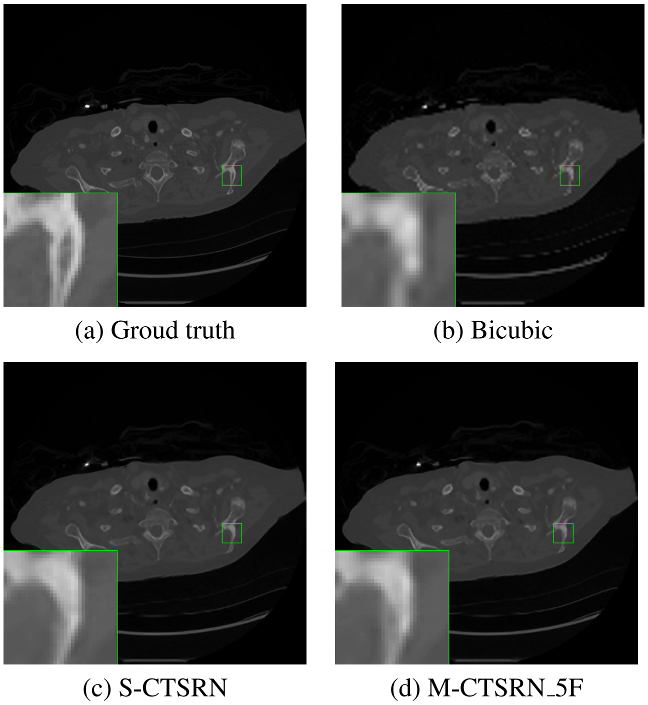
Currently, only a very limited number of CT datasets are available as research benchmarks. We manually collect over $10,000$ anonymized CT slices, and build our own dataset. We are considering the possibility to publicly release this CT dataset for research use.

In our experiments, we use a subset of $5,800$ slices to prepare HR-LR pairs as training input. A non-overlapping set of $1,000$ slices is reserved as test images.

4.2. Implementation Details

Both our single-slice and multi-slice models consist of two branches. The top residual branch is a three-layer CNN followed by a rearrange layer. The inputs are image patches of size 20×20 from the single slice or the multiple slices. The numbers of output feature maps of the three convolution layer are 64, 32, and s^2, where s is the super-resolving factor; their kernel sizes are set to be 5×5, 3×3, and 3×3, respectively. For the first two hidden layers, we use the ReLU activation. No zero-padding is applied in convolutional layers. The bottom upsampling branch uses a bilinear upsampling scheme to scale up the input slice by a factor of s, implemented by a deconvolutional layer with fixed kernel weights.

4.3. Results and Discussions



(a) Groud truth

(b) Bicubic

(c) S-CTSRN

(d) M-CTSRN_5F

Fig. 3. Qualitative results from 3 methods, (b) Bicubic, (c) S-CTSRN and (d) M-CTSRN, for x4 SR.

Scale	2	3	4
Bicubic	46.51	41.51	38.98
S-CTSRN	49.82	44.67	42.41
M-CTSRN_3F	50.04	44.91	42.64
M-CTSRN_5F	50.07	44.93	**42.81**
M-CTSRN_7F	**50.27**	**45.16**	42.73

Table 1. The average PSNR of different SR methods with SR scales of 2, 3 and 4. Best results are shown in bold.

In Tabs. 1 and 2, we compare S-CTSRN, M-CTSRN with different numbers of input slices, and the bicubic interpolation baseline at different SR scales. To evaluate the SR performance, we measure the Peak Signal-to-Noise Ratio (PSNR) and the Structural SIMilarity (SSIM) [30]. PSNR focuses on the pixel-based mean-squared error between the reconstructed images and ground truth ones, while SSIM accounts for the pixel covariances in local neighborhood between two images and usually reflects the image structural correspondences more faithfully.

As shown in Tabs. 1 and 2, S-CTSRN performs significantly better than the bicubic interpolation method quantitatively. M-CTSRN yields even better results than S-CTSRN,

Scale	2	3	4
Bicubic	0.98902	0.96965	0.95077
S-CTSRN	0.99292	0.98151	0.97082
M-CTSRN_3F	0.99322	0.98228	0.97237
M-CTSRN_5F	0.99326	0.98261	**0.97337**
M-CTSRN_7F	**0.99366**	**0.98341**	0.97325

Table 2. The average SSIM of different SR methods with SR scales of 2, 3 and 4. Best results are shown in bold.

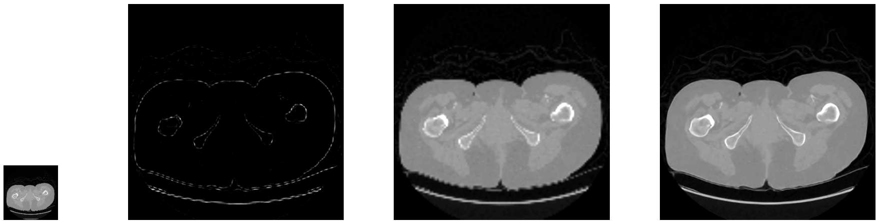

(a) LR input (b) Residual image (c) Bilinear upsampling (d) HR output

Fig. 4. Example input, intermediate and output images of our M-CTSRN_5F network for x4 SR.

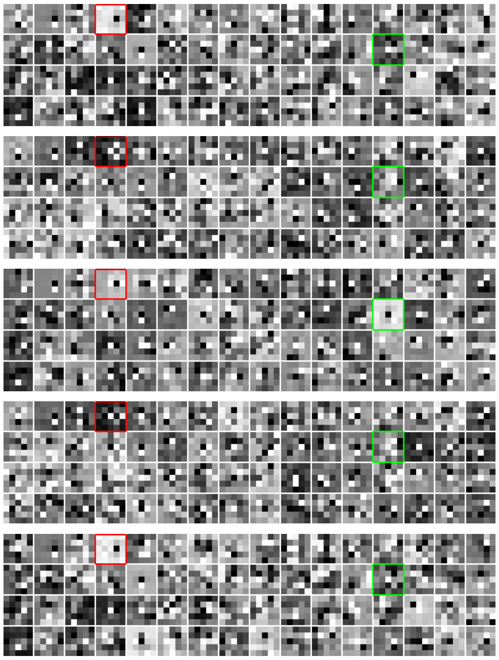


Fig. 5. Filters of the first convolution layer in M-CTSRN_5F model, each sub-image represents the learned filters from an input channel corresponding to one of the five input slices. For the sub-images in the red boxes, the first, third and fifth slices dominate in the SR process, meaning that the information from those slices contribute more to recover the missing information. For the ones in the green boxes, however, the central slice contributes the most for SR.

and shows the positive correlation between the improved quantitative results and the increased number of neighboring slices used as input. These conclusions can be further supported by visual comparisons shown in Figs. 3, 4, and 5. As shown in Fig. 3, S-CTSRN and M-CTSRN produce much more realistic details. Owing to the residual branch in our models, both proposed models explicitly target to learn the high-frequency details and to reconstruct the HR outputs with higher quality, which can be observed even more clearly from Fig. 4. The HR image produced directly by the upsampling branch has obvious aliasing artifacts along the edges. After adding the high-frequency residual component, the final HR output image becomes much sharper and the artifacts are notably suppressed.

Somewhat surprisingly, M-CTSRN seems to not only enhance the LR images, but even to compensate for some missing information during the degradation process. In the magnified green box area in Fig. 3, we can see that due to the original downsampling from ground truth, some intrinsic structure information about the dark ellipse area is missing at the center region, which cannot be recovered by either bicubic SR or S-CTSRN. By comparison, from the M-CTSRN results we are able to re-identify the missing dark ellipse structure again, thanks to the multi-slice coherence.

In Fig. 5, we show the learned convolution filters of the first layer. As can be seen, they tend to be primarily focused on corner points, which are usually crucial for the identifiability of fine structures in biomedical imagery. Compared to the filters learned for generic image SR [17], most of our filters have simpler structures, since CT images do not convey as much variable content as natural images.

Our method yields real-time SR due to the compact model size. The S-CTSRN and M-CTSRN super-resolve about 250 and 100 frames of 512×512 pixels per second using a single Nvidia Titan X Pascal GPU, respectively.

5. CONCLUSION

In the paper, we proposed single- and multi-slice CNN models for SR on CT images slices, which on one hand, provides the doctors with images of higher resolutions and therefore helps them better diagnose diseases, and on the other hand, protects patients from large dose of X-ray radiation. Our S-CTSRN model takes a single image as input and produces SR images, while our M-CTSRN model takes multiple consecutive CT slices, exploits their 3D correlation and produces the SR image for the central frame. Our experiments show that the both models achieve promising performance on a real-world CT dataset, while M-CTSRN yields the best results in terms of different evaluation measures.

6. REFERENCES

[1] A. C. Kak and M. Slaney, *Principles of computerized tomographic imaging*, SIAM, 2001.

[2] Z. Yan, J. Li, Y. Lu, H. Yan, and Y. Zhao, “Super resolution in ct,” *International Journal of Imaging Systems and Technology*, vol. 25, no. 1, pp. 92–101, 2015.

[3] D. L. Miglioretti et al., "The use of computed tomography in pediatrics and the associated radiation exposure and estimated cancer risk," *JAMA pediatrics*, vol. 167, no. 8, pp. 700–707, 2013.

[4] D. Poot, V. Van Meir, and J. Sijbers, "General and efficient super-resolution method for multi-slice mri," in *MICCAI*, 2010, pp. 615–622.

[5] W. Shi et al., "Cardiac image super-resolution with global correspondence using multi-atlas patchmatch," in *MICCAI*, 2013, pp. 9–16.

[6] F. Odille, A. Bustin, B. Chen, P. Vuissoz, and J. Felblinger, "Motion-corrected, super-resolution reconstruction for high-resolution 3d cardiac cine mri," in *MICCAI*, 2015, pp. 435–442.

[7] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *CVPR*, 2016, pp. 770–778.

[8] X. Wang, E. Türetken, F Fleuret, and P. Fua, "Tracking interacting objects using intertwined flows," *TPAMI*, vol. 38, no. 11, pp. 2312–2326, 2016.

[9] A. Maksai, X. Wang, and P. Fua, "What players do with the ball: A physically constrained interaction modeling," in *CVPR*, 2016, pp. 972–981.

[10] E. Türetken, X. Wang, C. Becker, C. Haubold, and P. Fua, "Network flow integer programming to track elliptical cells in time-lapse sequences," *TMI*, vol. 36, no. 4, pp. 942–951, 2017.

[11] S. D. Babacan, R. Molina, and A. K. Katsaggelos, "Variational bayesian super resolution," *TIP*, vol. 20, no. 4, pp. 984–999, 2011.

[12] J. Yang, J. Wright, T. S. Huang, and Y. Ma, "Image super-resolution via sparse representation," *TIP*, vol. 19, no. 11, pp. 2861–2873, 2010.

[13] Z. Wang, J. Yang, H. Zhang, Z. Wang, Y. Yang, D. Liu, and T. S. Huang, *Sparse Coding and its Applications in Computer Vision*, World Scientific, 2015.

[14] D. Glasner, S. Bagon, and M. Irani, "Super-resolution from a single image," in *ICCV*, 2009, pp. 349–356.

[15] Z. Wang, Y. Yang, Z. Wang, S. Chang, J. Yang, and T. S. Huang, "Learning super-resolution jointly from external and internal examples," *TIP*, vol. 24, no. 11, pp. 4359–4371, 2015.

[16] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," in *NIPS*, 2012, pp. 1097–1105.

[17] C. Dong, C. C. Loy, K. He, and X. Tang, "Learning a deep convolutional network for image super-resolution," in *ECCV*, 2014, pp. 184–199.

[18] Z. Wang, D. Liu, J. Yang, W. Han, and T. Huang, "Deep networks for image super-resolution with sparse prior," in *ICCV*, 2015, pp. 370–378.

[19] Z. Wang, Y. Yang, Z. Wang, S. Chang, W. Han, J. Yang, and T. Huang, "Self-tuned deep super resolution," in *CVPR Workshops*, 2015, pp. 1–8.

[20] D. Liu, Z. Wang, N.M. Nasrabadi, and T. S. Huang, "Learning a mixture of deep networks for single image super-resolution," in *ACCV*, 2016, pp. 145–156.

[21] A. Kappeler, S. Yoo, Q. Dai, and A. K. Katsaggelos, "Video super-resolution with convolutional neural networks," *IEEE Trans. on Computational Imaging*, vol. 2, no. 2, pp. 109–122, 2016.

[22] W. Shi, J. Caballero, F. Huszár, J. Totz, A. P. Aitken, R. Bishop, D. Rueckert, and Z. Wang, "Real-time single image and video super-resolution using an efficient sub-pixel convolutional neural network," in *CVPR*, 2016, pp. 1874–1883.

[23] A. Rueda, N. Malpica, and E. Romero, "Single-image super-resolution of brain mr images using overcomplete dictionaries," *Medical image analysis*, vol. 17, no. 1, pp. 113–132, 2013.

[24] E. Plenge, D. Poot, W. Niessen, and E. Meijering, "Super-resolution reconstruction using cross-scale self-similarity in multi-slice mri," in *MICCAI*, 2013, pp. 123–130.

[25] M. Protter, M. Elad, H. Takeda, and P. Milanfar, "Generalizing the nonlocal-means to super-resolution reconstruction," *TIP*, vol. 18, no. 1, pp. 36–51, 2009.

[26] O. Oktay et al., "Multi-input cardiac image super-resolution using convolutional neural networks," in *MICCAI*. Springer, 2016, pp. 246–254.

[27] Y. Zhang, G. Wu, P. Yap, Q. Feng, J. Lian, W. Chen, and D. Shen, "Reconstruction of super-resolution lung 4d-ct using patch-based sparse representation," in *CVPR*, 2012, pp. 925–931.

[28] D. Trinh, M. Luong, F. Dibos, J. Rocchisani, C. Pham, and T. Q. Nguyen, "Novel example-based method for super-resolution and denoising of medical images," *TIP*, vol. 23, no. 4, pp. 1882–1895, 2014.

[29] D. Liu, Z. Wang, B. Wen, J. Yang, W. Han, and T. S. Huang, "Robust single image super-resolution via deep networks with sparse prior," *TIP*, vol. 25, no. 7, pp. 3194–3207, 2016.

[30] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, "Image quality assessment: from error visibility to structural similarity," *TIP*, vol. 13, no. 4, pp. 600–612, 2004.