This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TMI.2018.2830381, IEEE Transactions on Medical Imaging

Enhancing the image quality via transferred deep residual learning of coarse PET sinograms

Xiang Hong[1], Yunlong Zan[2], Fenghua Weng[1], Weijie Tao[1], Qiyu Peng[3], Qiu Huang*[1,4]
[1] School of Biomedical Engineering, Shanghai Jiao Tong University, Shanghai, China 200240
[2] University of Michigan - Shanghai Jiao Tong University Joint Institute, Shanghai Jiao Tong University, Shanghai, China 200240
[3] Lawrence Berkeley National Laboratory, Berkeley, USA 94720
[4] Department of Nuclear Medicine, Ruijin Hospital, Shanghai, China 200240
*Email: qiuhuang@sjtu.edu.cn

***Abstract*—Increasing the image quality of positron emission tomography (PET) is an essential topic in the PET community. For instance, thin pixelated crystals have been used to provide high spatial resolution images but at the cost of sensitivity and manufacture expense. In this study, we proposed an approach to enhance the PET image resolution and noise property for PET scanners with large pixelated crystals. To address the problem of coarse blurred sinograms with large parallax errors associated with large crystals, we developed a data-driven, single-image super-resolution (SISR) method for sinograms, based on the novel deep residual convolutional neural network (CNN). Unlike the CNN-based SISR on natural images, periodically padded sinogram data and dedicated network architecture were used to make it more efficient for PET imaging. Moreover, we included the transfer learning scheme in the approach to process cases with poor labeling and small training data set. The approach was validated via analytically simulated data (with and without noise), Monte Carlo simulated data, and pre-clinical data. Using the proposed method, we could achieve comparable image resolution and better noise property with large crystals of bin sizes 4 times of thin crystals with a bin size from $1 \times 1\ mm^2$ to $1.6 \times 1.6\ mm^2$. Our approach uses external PET data as the prior knowledge for training and does not require additional information during inference. Meanwhile, the method can be added into the normal PET imaging framework seamlessly, thus potentially finds its application in designing low cost high performance PET systems.**

***Index Terms*—positron emission tomography, super resolution, sinogram, deep residual learning, convolutional neural networks, transfer learning**

I. Introduction

POSITRON emission tomography (PET) is an efficient molecular imaging technique in functional imaging, providing information on perfusion, metabolism or receptor binding. PET has been widely used in neurology, oncology and cardiology [1]–[3]. It is considered highly sensitive in diagnosis because of its ability to detect lesions that are not visible for anatomical imaging modalities such as computed tomography (CT) or magnetic resonance imaging (MRI). However, PET imaging is limited by its low image quality, especially low spatial resolution and high noise level, compared with CT and MRI.

Extensive attempts have been exerted to improve the image quality of PET systems. For instance, small pixelated crystal sensors and the depth of interaction (DOI) technology were used to improve the spatial resolution [4]–[6] thanks to their accurate positioning of incident photons. However, DOI requires complicated detector designs. And thin crystal sticks are difficult to manufacture and hence expensive. Meanwhile, in a crystal array built with thinner crystals, the dead space caused by the inter-crystal filling is generally larger, which decreases the detection efficiency of pixelated crystal detectors. The statistical error is then larger for these PET systems given a fixed dose. To better utilize the resolution-efficiency tradeoff, PET systems with monolithic crystals [7] were proposed. However, when multiple silicon photomultipliers are mounted to one monolithic crystal, extra efforts are required on estimating positions of incident gamma photons. The best estimation accuracy was reported to be less than 2 mm in the literature, but with a few assumptions such as the short crystal length and limited incident angles of photons [8]–[10]. Another category of approaches to the high image quality of PET systems is via image processing. Conventionally, applying regularizations and adapting the hyper parameter make a tradeoff between the signal to noise ratio (SNR) and the image resolution [11]. Modeling and correcting for the point spread function (PSF) finds another way to improve resolution and uniformity [12], [13]. Others used super-resolution (SR) methods. SR is widely used in many fields for enhancing the resolution of sensing systems. Many different SR methods have been published for PET imaging. Some required supernumerary data with temporal or spatial coherence [14] , and others used joint methods of SR with multi-modality data [15]. These methods can achieve impressive performance, but they often require extra information or even a registration process that may introduce extra noise.

In this work we proposed an approach to improve the image SNR and resolution for PET scanners with large pixelated crystals. By using large pixelated crystals we avoided the extra efforts to estimate incident photon positions. We only targeted the coarse blurred sinogram and large parallax errors, using a

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TMI.2018.2830381, IEEE Transactions on Medical Imaging

data-driven, single-image super-resolution (SISR) method.

There are several types of SISR methods that focus on learning mapping from low-resolution images to high-resolution ones. Two popular categories of methods among them are sparse-coding-based methods and deep convolutional neural network (CNN)-based methods.

The sparse-coding-based SISR has been extensively studied in the computer vision field community. The current trend is via external example-based SISR methods [16]–[19]. These methods obtain low-resolution to high-resolution mapping by learning dictionaries or manifold spaces corresponding to low-resolution patches and high-resolution patches. Applications can also be found on medical image super-resolution studies [20]–[22] . However, due to the high computational cost, many of these methods use relatively small patches, usually smaller than ten pixels per dimension. Using small patches limits the receptive field and performance of these methods. Furthermore, sparse-coding-based methods often need carefully designed filters for feature extraction, requiring much effort from experts. Besides, during the inference process, some sparse-coding-based methods need large physical storage and a long search latency.

Deep CNN-based SISR is developed after deep CNNs draw much attention due to their impressive performance on image classification [23]–[25] . Deep CNNs have also been successfully applied to many different fields such as object detection [26]–[28] and medical image processing [29]–[32]. As shown in [33], sparse-coding-based methods can be viewed as a special case of deep CNNs with smaller receptive fields and limited operations to optimize. On the other hand, CNNs are designed to solve the problem of obtaining features via hand crafted convolution kernels. Thus CNNs learn representations (features) of images automatically. Also the inference process is fast in deep CNNs since it is independent of the size of the training set and only forward propagation is required. In general, deep CNNs outperform sparse-code-based methods, especially for natural image super resolution tasks and in the computer vision community, where deep CNNs directly learn end-to-end mapping between low-resolution images and high-resolution images [33], [34]. However applying deep CNNs directly to medical image processing may lead to poor performance. Firstly, commonly used data augmentation techniques on natural images are not valid for medical images. Secondly, medical images require a higher accuracy in the intensity, compared with natural images. Hence, the network architecture of deep CNNs for medical images needs to be carefully designed to avoid the noise.

We used the deep residual learning to perform super resolution on PET sinograms as in [35]. However the network architecture was redesigned so that the image quality could be improved even with a higher down-sampling ratio. The CNN-based method was designed and applied as an intermediate step between the projection data acquisition and the image reconstruction. Transfer learning was used to boost performance of proposed method as shown in section II. Then in the section followed, we validated the method with both analytically simulated data, Monte Carlo (MC) simulated data and experimental pre-clinical data. Finally we discussed the CNN design issues and concluded the study.

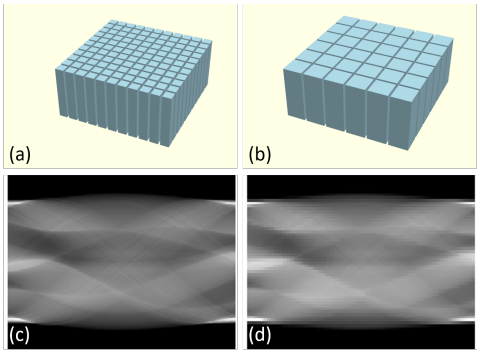

Fig. 1. Detector and sinogram for thin crystals (a) (c) and large crystals (b) (d).

II. Method

A. Overview

The commercialized whole-body PET scanners are usually built with pixelated crystals of a bin size about $4 \times 4\ mm^2$. Recently the crystal size of $2.35 \times 2.35\ mm^2$ was reported by some vender [36] . In this work, we validated our method on simulated scanners with several bin sizes from $2 \times\ 2\ mm^2$ to $12.8 \times 12.8\ mm^2$, noted as the large crystals. Large crystals were compared with thin crystals with a bin size of $1 \times 1\ mm^2$ to $1.6 \times 1.6\ mm^2$. Fig. 1 shows the thin pixelated crystal array and the large pixelated crystal array. As mentioned above, the array made of large crystals provides higher detection efficiency than that with thin crystals. However, as Fig. 1 illustrates, the sinogram for the large crystal array is blurred due to the coarse sampling, and large parallax errors are also introduced. To solve these problems, we used a deep learning-based, single-image super-resolution (SISR) method.

Our proposed algorithm generated the inference super-resolution sinogram (SRS) by performing SISR on the low-resolution sinogram (LRS) obtained with large crystals. The process required prior information learned from external PET image data, a.k.a., the high-resolution sinogram (HRS) simulated to mimic the acquisition with thin crystals. Three networks were trained for large crystals with bin sizes 2, 4 and 8 times those of thin crystals, called networks with $2\times$, $4\times$ and $8\times$ down-sampling, respectively. Note that the networks with same down-sampling ratio may have different weights for different tasks. For simplicity, we used the term $x\ mm$ crystal system to indicate systems with crystal of size $x \times x\ mm^2$.

Once the network had been trained, the SRS, together with LRS, HRS, and BIS (the bilinear interpolation of LRS), was reconstructed to obtain four images, as shown in Fig. 2 (a). We compared the low-resolution sinogram reconstructed image (LRI), the interpolated sinogram reconstructed image (IRI), the super-resolution sinogram reconstructed image (SRI), and the high-resolution sinogram reconstructed image (HRI). We also compared the performance of applying our network on the projection domain and on the image domain, where the procedure for the latter is shown in Fig. 2 (b). The network for image domain is trained using reconstructed images from same

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TMI.2018.2830381, IEEE Transactions on Medical Imaging

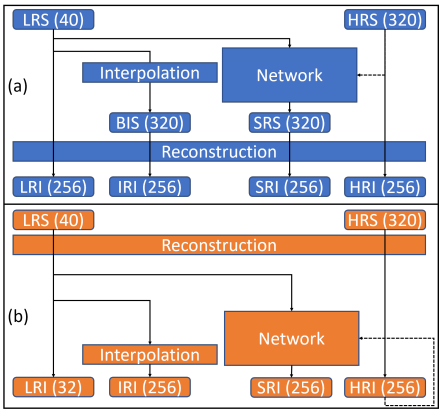

Fig. 2. Procedures of applying our network on sinograms (a) and on reconstructed images (b). Numbers in parentheses indicate sizes of corresponding data in an example of $8\times$ down sampling, e.g. "LRS (40)" means low resolution sinogram of size 40×40. LRS, BIS, SRS, HRS are sinograms of low resolution, bilinear interpolation, network-based super-resolution and high-resolution, while LRI, IRI, SRI and HRI are images reconstructed from these sinograms. The procedure presented with dashed line is only present in the training phase.

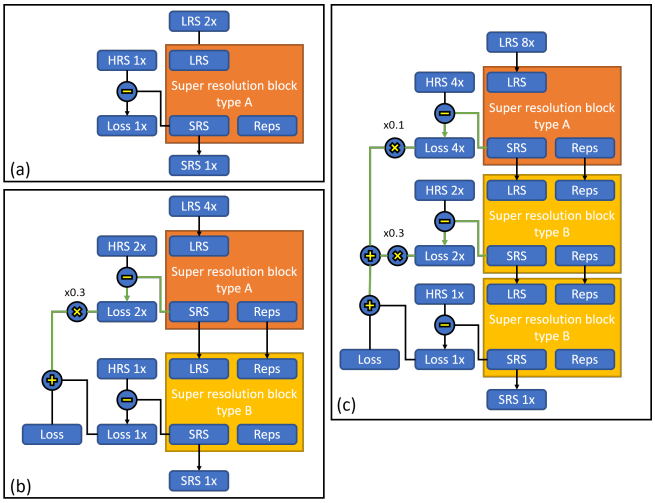

Fig. 3. Networks for down-sampling ratios of 2(a), 4(b) and 8(c). Details of super-resolution block are shown in Fig. 4, where "Reps" (short for "representations") is a feature tensor with 64 channels as one of outputs of residual blocks. Additional short-cut paths for multiscale loss are colored in green. For simplicity, crop layers are not shown here.

sinograms used for training network for projection domain. When applying the network directly on reconstructed images, there is no BIS, SRS, IRI or SRI. However for simplicity, we keep these terms.

B. Formulation

In the SISR task, we were looking for the optimal end-to-end mapping from low-resolution sinograms to high-resolution sinograms, which can be formulated as follows:

$$f^* = \arg\max_f \Pi_{(x,y)\in D} P(y|f(x)), \tag{1}$$

where $x \in R^M$ is the LRS, $y \in R^N$ is the HRS, and $f(x) \in R^N$ is the SRS with $N > M$. (x, y)represents a pair of LRS and HRS from external dataset D. We simply choose $P(y|f(x))$ to be Gaussian. Thus, the optimal mapping f^* can be achieved by solving a maximum likelihood problem:

$$f^* = \arg\min_f \sum_{(x,y)\in D} ||y - f(x)||^2, \tag{2}$$

since x and y share much information, and it is inefficient to learn the mapping $f(\cdot)$ from x to y directly, as suggested in [34] , we reformulate (2) into

$$f^* = \arg\min_f \sum_{(x,y\in D)} ||y - (b(x) + r(x))||^2, \tag{3}$$

where $b(x)$ is bilinear interpolation of x, and $r(x) = f(x) - b(x)$, is the residual between SRS $f(x)$ and the BIS $b(x)$. Three types of deep residual sinogram super-resolution networks (DRSSRNs) were then trained according to Eqn. (3).

C. Deep residual sinogram super-resolution network (DRSSRN)

Our DRSSRN is constructed by several super-resolution blocks (SRBs) as shown in Fig. 3 . The number of SRBs in DRSSRN varies according to the down-sampling ratio. We used 1, 2 and 3 SRBs for the DRSSRN with down-sampling ratio of 2, 4 and 8, respectively. Each SRB works as a sub-super-resolution network with an up-sampling ratio of 2.

HRS and SRS were passed to crop layers before calculating the loss. The number of pixels cropped from each edge were 4, 8 and 16 for $8\times$, $4\times$ and $2\times$ down-sampling respectively. For DRSSRNs with down-sampling ratios of 4 and 8, we added short cut paths (in green lines) as shown in Figs. 3 (b) and (c) to aid in gradient back propagation. We used output representations of the inference block to reconstruct the middle results. Training the network was faster when we added sinograms with down-sampling ratios of 2 and 4 as middle results and added middle losses with weights of 0.3 and 0.1 respectively to fit different scales of multiscale loss.

Details about SRB are shown in Fig. 4. There are two types of super-resolution blocks. Type A is used in the input of the DRSSRN and consists of a low-order representation extraction block to obtain representations as input for the super resolution block. A bilinear interpolation layer followed by a convolutional layer was used to form a low-order representation extraction block in our work. SRB of type B uses interpolated representation of previous SRB.

In the reconstruction block, we used a convolution layer with one output channel to obtain the reconstructed residual intensity and obtained SRS by adding this residual to the interpolated LRS.

The inference block is a sequential stack of 20 residual blocks shown in Fig. 5. The residual block is similar to [25] with bottleneck layer replaced by inception block [24]. The use of an inception block instead of a stack of plain CNNs increases the receptive field and model capacity, while ensuring the number of variables is not increasing too fast.

We used the concatenate exponential linear unit (CELU) as the activation function $CELU(z) : R \to R^2$, which can be

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TMI.2018.2830381, IEEE Transactions on Medical Imaging

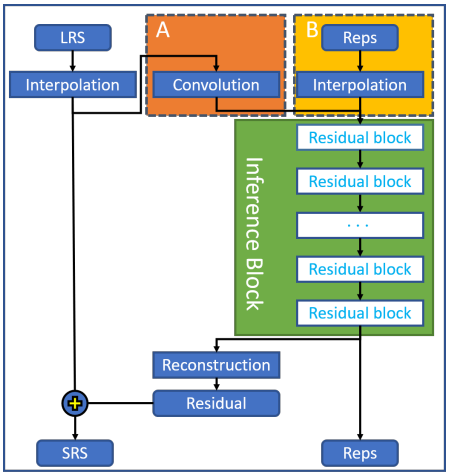

Fig. 4. Details of the super-resolution block. The upper right corner shows different versions of type A and type B.

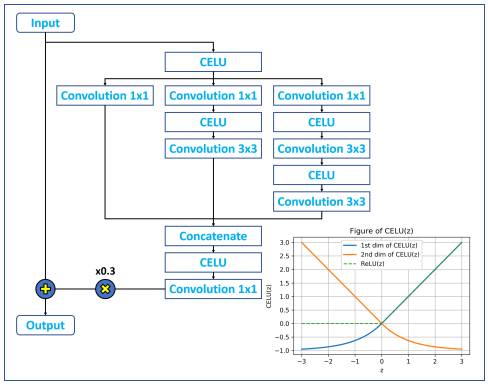

Fig. 5. Details of the residual block. A scaling coefficient $\alpha = 0.3$ is used on outputs of the inception block to make the training process faster and more stable. Activation function CELU and conventional ReLU are plotted on the bottom right.

formulated as follows:

$$CELU(z) = (ELU(z), ELU(-z))^T, \tag{4}$$

where $ELU(\cdot)$ is the exponential linear unit (ELU) [37] that can be formulated as follows:

$$ELU(z) = \begin{cases} z, & z > 0 \\ \exp(z) - 1, & z \leq 0 \end{cases} \tag{5}$$

Two dimensions of output of CELU, as well as ReLU for comparison, are shown in right bottom of Fig. 5.

In practice, we used overlapping patches $\{P_i^k\}$ cropped from low-resolution sinogram x_i as input to DRSSRN instead of the whole sinogram, reducing the time and memory costs for both training and inference tasks. Before passing these patches into the network, we normalized each patch via $\tilde{P}_i = \frac{P_i - mean(x_i)}{std(x_i)}$.

During the training phase, we mixed and shuffled patches from different sinograms to construct mini-batches to fully utilize computational resources of graphical processing units (GPUs). We implemented the DRSSRN using the Tensorflow [38] framework.

In addition, we performed periodical padding on the angle axis and zero padding on the sensor number axis for both training and testing. During training, this padding can be regarded as data augmentation and makes our network invariant to rotation transformation. During testing, padding is required to handle cropped pixels at the boundary.

All networks were trained with Adam optimizer [39] with default configurations except that the initial learning rate was set to 0.0001.

The DRSSRN performance can be potentially increased if we used more residual blocks in inference block to increase the network capacity and more training data to avoid over-fitting. However it is difficult to obtain a lot samples via Monte Carlo simulations or experiments. On the other hand, the network performance heavily depends on quality of labels. We used high resolution sinograms as labels in this case. With the high noise present in PET data, we had very poor labeling, which made the result of DRSSRN even worse than interpolation. We use transfer learning to tackle problems mentioned above. The idea is to use network trained in "easy" tasks as initial value for training network for "complicated" tasks. For example, we started with weights of fully trained DRSSRN on clean analytically simulated data, and tuned the network for noisy analytically simulated data, which resulted in much higher performance than the network trained from scratch. Transfer learning was also used for Monte Carlo simulated data and pre-clinical data.

III. Analytic Simulation

The DRSSRN was first trained and evaluated using simulated data with four commonly used digital phantoms.

The first three are randomly deformed from the Shepp-logan, Derenzo, and Jaszczak phantoms. Their sinograms are called "Analytically simulated Phantom Sinogram (APS)" in the rest of the article. The more realistic XCAT [40] phantoms with defects are used especially to assess the lesion detectability.

A. Digital phantoms

Shepp-logan-like phantoms were constructed using 10 ellipses, each with 6 parameters (major axis, minor axis, rotation, positions and intensity). To generate a phantom, we randomly sampled these parameters from the normal distribution $N(\mu, 0.1\mu)$, where μ is the original value of each parameter in the standard phantom.

For Derenzo-like phantoms, we uniformly sampled the number of spots of each edge along triangles in six sections according to a distribution $U(1, 19)$. Two intensity values were sampled randomly according to $U(0.0, 1.0)$ for the background and hot spots, respectively, with the condition that background has lower intensity. Here $U(min, max)$ indicates the uniform distribution with minimum and maximum values.

Similarly, the number of circles in Jaszczak-like phantoms was first chosen randomly according to $U(6, 10)$. Then the radius and intensity for each circle were decided according to $U(0.0, 1.0)$.

XCAT phantoms were generated with a cardiac defect of different severity.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TMI.2018.2830381, IEEE Transactions on Medical Imaging

Fig. 6. Examples of analytically simulated phantoms.

Fig. 6 shows a few examples of the 786543 Shepp-logan-like, Derenzo-like and Jaszczak-like phantoms, together with 4 XCAT phantoms with no defect, 25%, 37.5% and 50% of defect.

B. Analytically simulated sinograms

High-resolution sinograms (320×320) were generated to simulate thin crystals with a bin size of $1 \times 1\ mm^2$. We used summation pooling with windows 2×2, 4×4 and 8×8 on high-resolution sinograms to simulate sinograms from large crystal systems with bin sizes of $2 \times 2\ mm^2$, $4 \times 4\ mm^2$, $8 \times 8\ mm^2$. In order to mimic different dose levels, we also normalized summation of sinograms to $1e6(100\%)$ and generated samples with Poisson noise. For XCAT phantoms, data were simulated for even smaller dosages of $1e5(10\%)$, $5e4(5\%)$, $2.5e4(2.5\%)$.

C. Training of DRSSRN

In total 80% phantoms in each category were used for training and the other 20% for testing. Three DRSSRNs were trained from scratch using clean $2\times$, $4\times$ and $8\times$ down sampling sinograms. Then they were used as initial parameters for training of noisy sinograms. During training we used 256×256 patches from data, thus inputs for $2\times$ and $8\times$ down sampling networks were 128×128 and 32×32 respectively, with mini-batch size equal to 6. DRSSRN was trained for $3.2e5$ iterations for clean data. While for noisy data, we trained additional $1.0e5$ iterations based on network initialized with weights trained on clean data.

Since training or fine-tuning with only 4 XCAT phantoms is not applicable, the DRSSRN trained by noisy APSs were used directly for tests on XCAT phantoms. Similar training and testing were performed on reconstructed images for comparison.

D. Metrics

We assessed the DRSSRN on 500 Shepp-logan-like, Derenzo-like, and Jaszczak-like phantoms in the test set in terms of the resolution, multi-scale structural similarity index (MS-SSIM) [41], and peak signal-to-noise ratio (PSNR). PSNR is defined as follows:

$$PSNR = \frac{10}{\log 10} \log \frac{255^2}{\sum_{ij}^{MN} (P_{ij} - I_{ij})^2}, \tag{6}$$

where P_{ij} is the reference image of size $M \times N$ and I_{ij} is the image to be measured. The image intensity was rescaled to range $[0, 255]$. Besides, the contrast recovery coefficients (CRC) was calculated on 50 Derenzo phantoms as follows:

$$CRC = (\frac{I_{hot}}{I_{cold}} - 1)/(\frac{P_{hot}}{P_{cold}} - 1) \tag{7}$$

where I_{hot} and I_{cold} are mean intensities of hot and cold regions of the image, and P_{hot} and P_{cold} are for the reference image.

To assess the noisy property, we studied the bias and variance tradeoff of images reconstructed from noisy sinograms. By randomly selecting 32 phantoms and generating 100 noisy sinograms for each phantom, we calculated the bias and variance for each voxel of a specific phantom. Then the mean bias and the mean variance of the phantom were plotted.

Finally the lesion detectability was evaluated using XCAT phantoms, via the Channelized Hotelling Observer (CHO) [42].

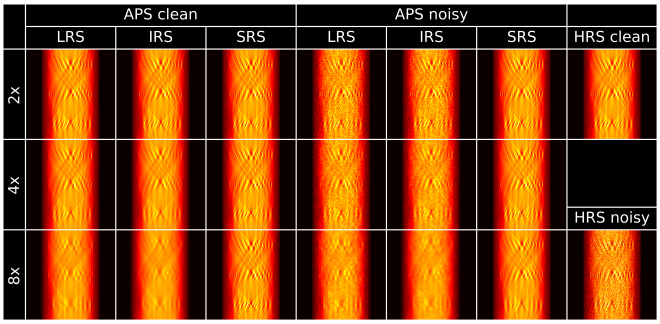

Fig. 7. Sinograms of different processing methods.

E. Results

Fig. 7 shows sinograms obtained by applying the DRSSRN. The effect of de-noising is evident for SRS. Using these sinograms images were reconstructed via filtered back projection (FBP) algorithm, as shown in Fig. 8. The result obtained by applying the same network on FBP reconstructed images is also shown. The HRI was reconstructed from sinograms of $1\ mm$ crystals. When noise-free data are concerned, the DRSSRN-processed sinogram (the 3rd column) for $2\ mm$ and $4\ mm$ large crystals can achieve acceptable image quality, compared with HRI. Even for $8\ mm$ large crystal system, SRIs can recover more details than direct reconstruction (the 1st column) of $2\ mm$ crystals. Similar trend can be observed with noisy (with totally $1e6$ events for each image) data. DRSSRN works better in the projection domain than in the image domain for all the down-sampling ratios, no matter if noise is present in sinograms.

A comparison of profiles is shown in Fig. 9, for noisy test data in the popular $4\ mm$ crystal system. The comparison was made among profiles of the phantom (the ground truth), HRI, LRI, IRI and SRI of $4\times$ down-sampling ($4\ mm$ crystal system) for noisy test data. Noise in SRI is much lower than that in HRI and LRI. The boundary of SRI is sharper and more accurate than that of IRI. This agrees with the visual observation in Fig. 8 and is also illustrated in the quantitative comparisons in Fig. 10.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TMI.2018.2830381, IEEE Transactions on Medical Imaging

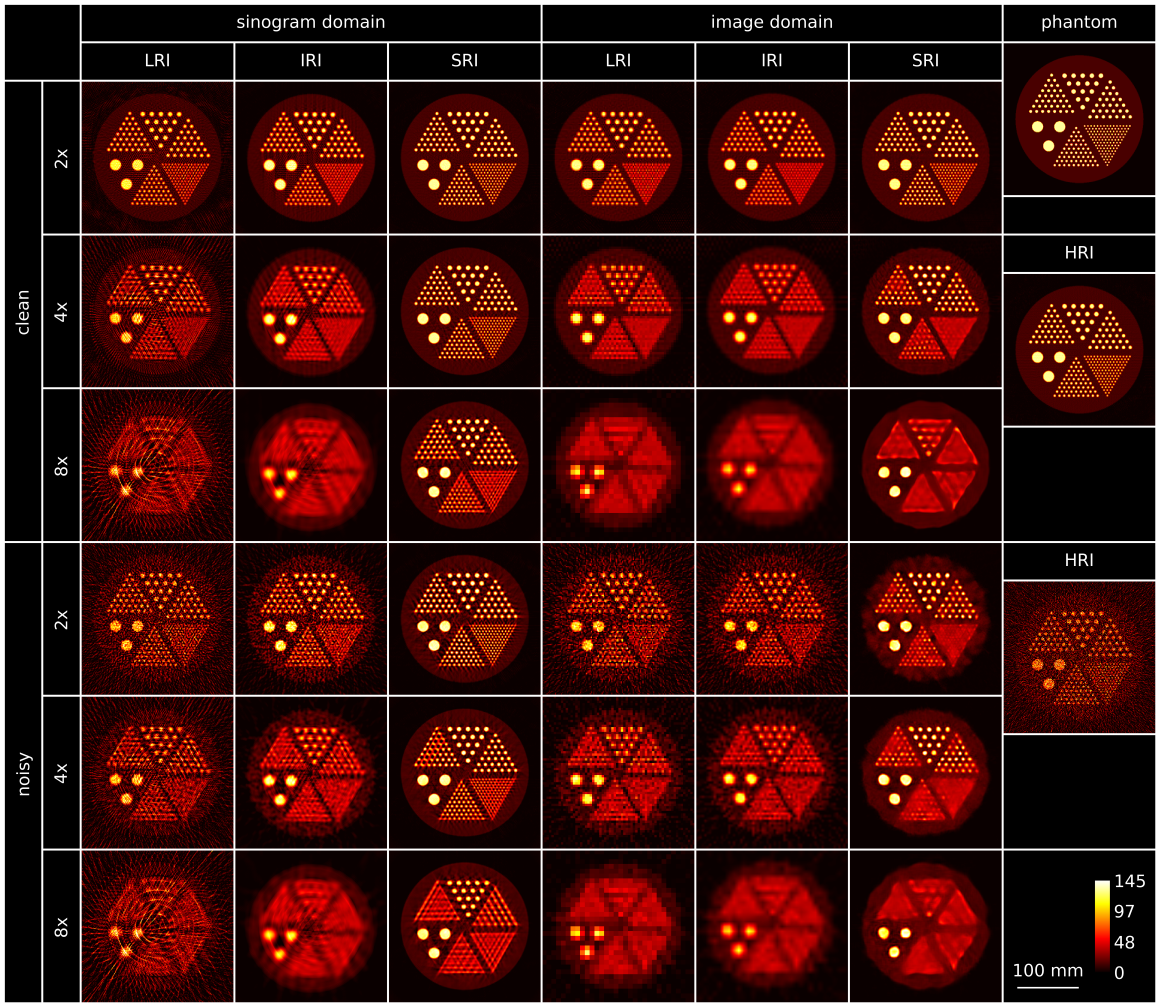

Fig. 8. Images obtained using our network both on sinograms and on reconstructed images. The proposed network works better on sinograms than on images, for both clean and noisy data. SRI outperforms LRI and IRI, no matter if the network is applied in the sinogram domain or in the image domain. SRI of our network on sinograms with $2\times$ and $4\times$ down-samplings achieves the best image quality among all.

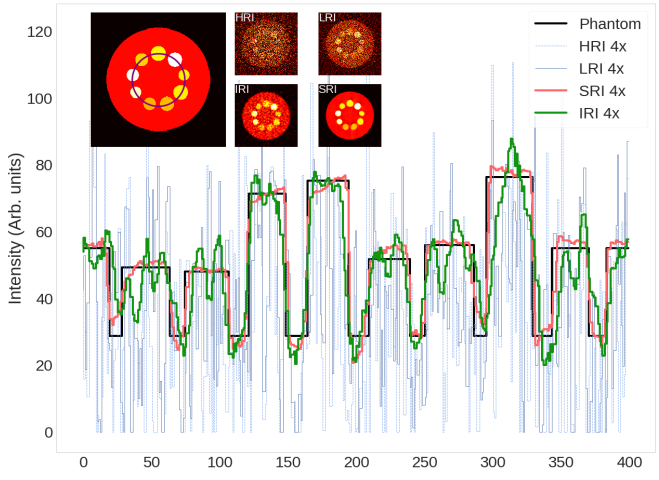

Fig. 9. Profiles of one noisy data test with $1e6$ events with $4\times$ down-sampling. SRI shows less noise, sharper and more accurate edges than others.

As shown in Fig. 10, images with super-resolution sinograms have higher PSNR, MS-SSIM and CRC than those from direct reconstruction (low resolution) or interpolated sinograms with the same crystal bin size, no matter if there is noise. With noise free data, the image quality of SRI becomes worse with the down-sampling ratio increases. However noise hampers PSNR and MS-SSIM, with more severe degradations for the thin crystal system than for the large crystal system. Thus with noisy data, HRI performs worse than all others, and SRI $4\times$ and SRI $2\times$ show comparable performance. SRIs of all down-sampling ratios also show better CRC than HRI.

The bias and variance tradeoff shown in Fig. 11 illustrates that SISR with DRSSRN achieves better performance than other processing methods of the same crystal size. SRI $2\times$ and $4\times$ both have less bias and variance than other processing methods of all crystal sizes.

Fig. 12 presents reconstructed images of the XCAT phantom study. As expected, in the LRIs, higher noise is observed with lower dose. Both IRI and SRI show their ability of de-noising, but SRI provides images with sharper edges and higher contrast than IRI. Even for a down-sampling ratio of 8, SRI has no streak artifacts. When there is a defect in the XCAT phantom, SISR with DRSSRN provides better or at

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TMI.2018.2830381, IEEE Transactions on Medical Imaging

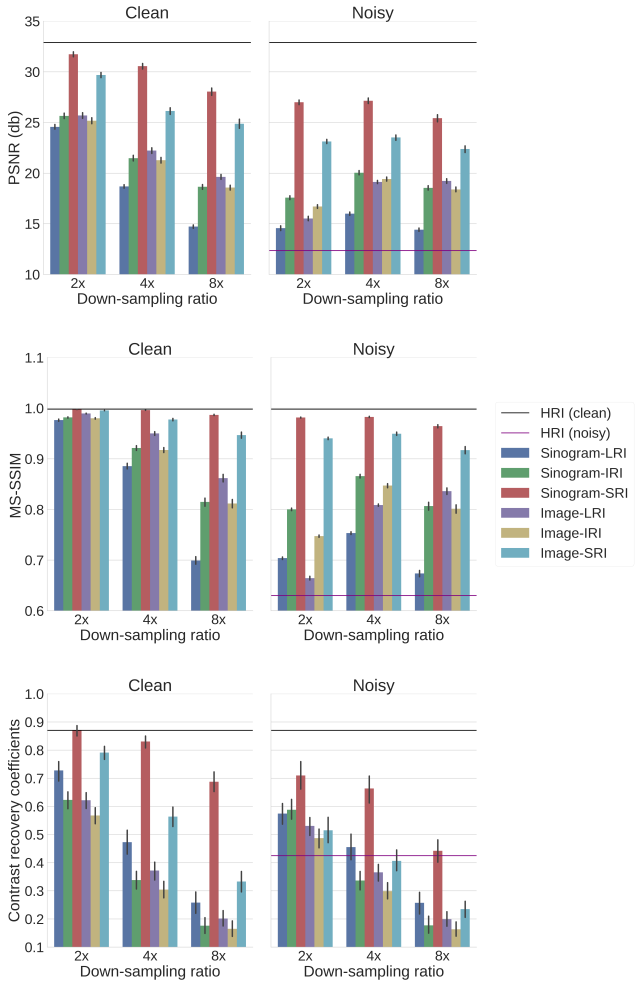

Fig. 10. Metrics for different processing methods, down-sampling ratios, and noise levels. For all cases with fixed down-sampling ratio and noise level, SRI is superior to LRI and IRI in terms of PSNR, MS-SSIM and CRC. For noisy cases, SRI even outperforms HRI. SRI with $2\times$ and $4\times$ down-samplings achieves comparable performance as clean HRI.

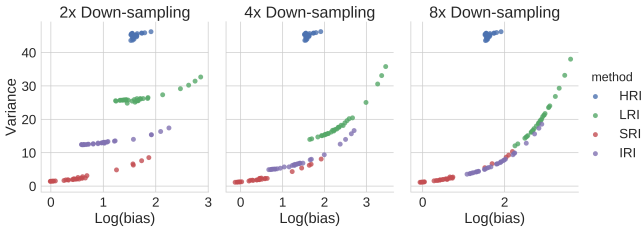

Fig. 11. Bias and variance tradeoff for DRSSRN with $1e6$ events. For any specific processing method, bias increases and variance decreases with down-sampling ratio increasing. SISR of sinogram with DRSSRN get minimum average bias and variance in all three tested down- sampling ratios.

least comparable detectability, compared to other processing methods including reconstructed image from high resolution sinograms, as shown in the receiver operating characteristic (ROC) curves in Fig. 13 (a). The area under curves (AUCs) are calculated and then the differences in AUC of SISR to other methods are shown in Fig. 13 (b), with different dose levels, defect intensities, and down-sampling ratios. The maximum

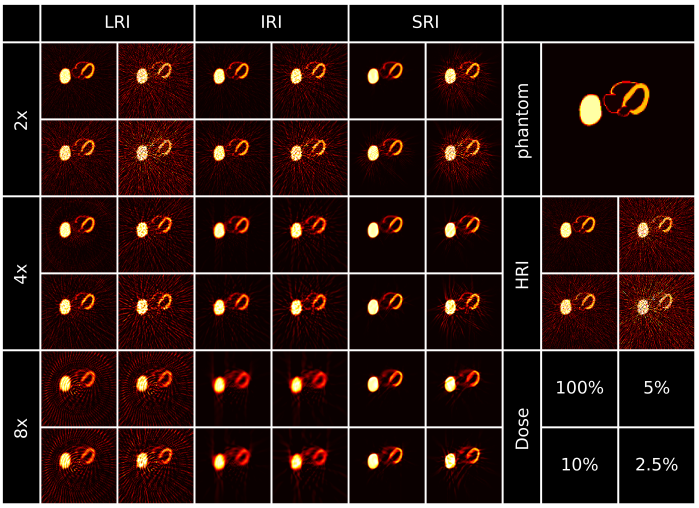

Fig. 12. Reconstructed XCAT data. SISR with DRSSRN is less sensitive to noise associated with low dose.

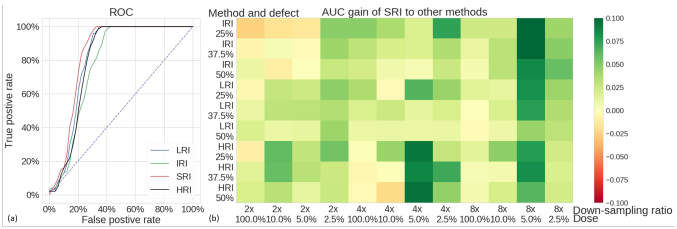

Fig. 13. ROC of 50% defect, 2.5% dose and $8\times$ down-sampling (a) and AUC gains for different methods (b). SISR with DRSSRN improves lesion detection.

gain of using SISR is 0.13. Some negative gain is observed here, with the largest absolute value being 0.02.

IV. Monte Carlo Simulation

A. Data generation

We performed Monte Carlo (MC) simulations of PET scans in GATE [43] for a series of mCT-like [44] scanners. To generate sinograms of different resolutions, we used four settings of detectors each with one detector ring: 640 $3.2\times3.2\times20\ mm^3$ crystals, 320 $6.4\times6.4\times20\ mm^3$ crystals, 160 $12.8\times12.8\times20\ mm^3$ crystals and 80 $25.6\times25.6\times20\ mm^3$ crystals. Activity and time of acquisition were set to obtain roughly $1e6$ events per scan. Scattering and attenuation in phantom were not considered in order to accelerate simulations. We generated 4 sinograms for each of the 983 randomly picked phantoms from the previous section. In total 80% of these data (sinograms from 786 phantoms) were used as training dataset, and the rest for testing.

It is hard to train the DRSSRN for MC simulated data from scratch, due to the poor labeling from noisy sinograms. In the study we found that SRI via training from scratch did not outperform IRI. Hence we used transfer learning with the DRSSRN trained by noisy APSs as the initialization. MC simulated data were only used for fine-tuning. Training parameters were similar to those used in III-C, with the iteration number equal to $7.1e4$.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TMI.2018.2830381, IEEE Transactions on Medical Imaging

TABLE I
RESOLUTION OF IMAGES FROM DIFFERENT TYPE OF SINOGRAMS AND CRYSTAL BIN SIZE OF MC SIMULATED DATA

Crystal bin size	HRI & LRI	IRI	SRI
1.6 mm	2.22 mm	N/A	N/A
3.2 mm	3.25 mm	3.16 mm	2.83 mm
6.4 mm	6.56 mm	6.56 mm	2.25 mm
12.8 mm	16.19 mm	16.19 mm	4.24 mm

B. Results

Reconstructed images in Fig. 14 shows up to $4\times$ down-sampling, SRI has higher resolution and less noise than LRI and IRI. We obtained the point spread function of tiny hot spots in Derenzo phantoms and fit with a Gaussian function to measure resolution, as in Table I. SRI $4\times$ (i.e., $6.4\ mm$ crystals) is capable of improving the resolution to match the HRI ($1.6\ mm$ crystals). PSNR and MS-SSIM in Fig. 15 show the same trend as in the analytic simulations. However the CRC for SRI is not as good as HRI. It is probably due to the high variance in CRC calculation caused by small number of Derenzo phantoms. Note that SRI achieves the best performance with $4\times$ down-sampling, which agrees with the observation in analytic simulations.

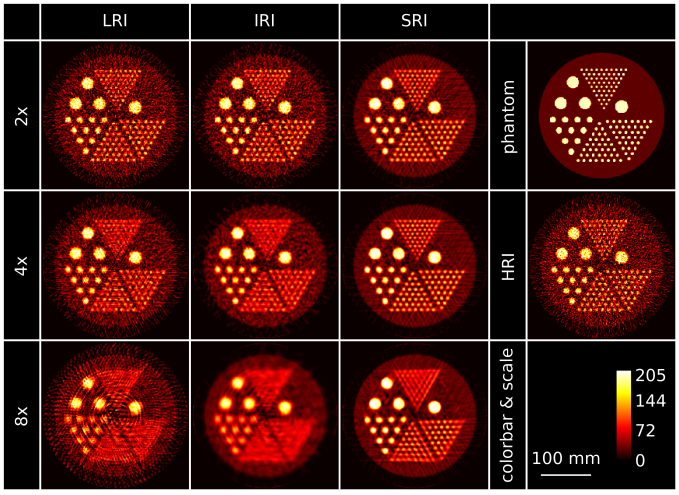

Fig. 14. Reconstructed MC simulated data. Up to $4\times$ down-sampling, most hot spots are separable on SRI with less noise, compared to LRI and IRI.

V. PRE-CLINICAL STUDY

A. Data acquisition

Data of 30 normal mice were also acquired to evaluate our algorithm, among which 80%, i.e. 24 mice were used for training and the rest 6 for testing. Each mouse was scanned for 5 to 30 minutes using Siemens Inveon preclinical system with an injection of 0.2 mCi ^{18}F-FDG. The detector bin size of the Inveon PET system is $1.59 \times 1.59 \times 10\ mm^3$. The list-mode data from Inveon PET system were arranged to oblique sinograms with a span of 3 and a max ring difference of

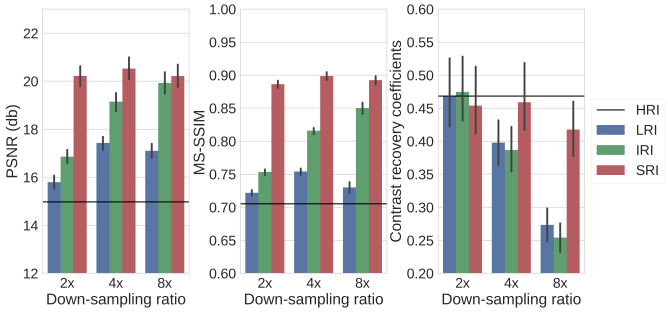

Fig. 15. Metrics for MC simulated data. PSNR and MS-SSIM of SRI images are significantly higher than those for LRI and IRI. CRC does not show the same trend. It may be due to the high standard deviation of CRC. SRI achieves the best performance with $4\times$ down-sampling for different ratios.

79. Then, the oblique sinograms were rebinned into 159 2D-sinograms (128×320) using a Fourier rebinning algorithm [45].

We used these rebinned sinograms as reference sinograms. Summation pooling with window sizes of 2×2, 4×4 and 8×8 were used on the rebinned sinograms to mimic large crystals with bin sizes of $3.18 \times 3.18\ mm^2$, $6.36 \times 6.36\ mm^2$ and $12.72 \times 12.72\ mm^2$. Considering these data as high dose scans, we added Poisson noise to them to evaluate our SISR method on low dose cases. In particular, we generated noisy mice data by generating random number under the Poisson distribution with means equal to 10% of original data. We used 128×64 patches for training DRSSRN with $1.2e5$ iterations.

B. Results

For mice data, reconstructed images via MLEM reconstruction are shown in Fig. 16, and zoomed transverse images are shown in Fig. 17. The quality of HRI, LRI and IRI degraded quickly with crystal bin size increasing, while SRI shows negligible changes. The image quality of SRI is tolerable even with $12.72\ mm$ large crystals. In the low dose case, SRI also shows ability of de-noising. We did not perform quantitative comparisons since there is no ground truth in this study.

VI. DISCUSSION

In this study, we developed a data-driven, single-image super-resolution method on sinograms of large crystal systems. Some strategies different from conventional CNN-based SISR approaches were used in our DRSSRN to improve performance.

A. Architecture of the network

Inspired by the work in [34] that presented a CNN based SISR model, we learned the residual between the high-resolution sinogram and low-resolution sinogram, instead of learning the high-resolution sinogram itself, to accelerate the training of DRSSRN.

Since the receptive field of pixels near a patch boundary exceeds the image boundary, crop layers were added before calculating the loss in the network.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TMI.2018.2830381, IEEE Transactions on Medical Imaging

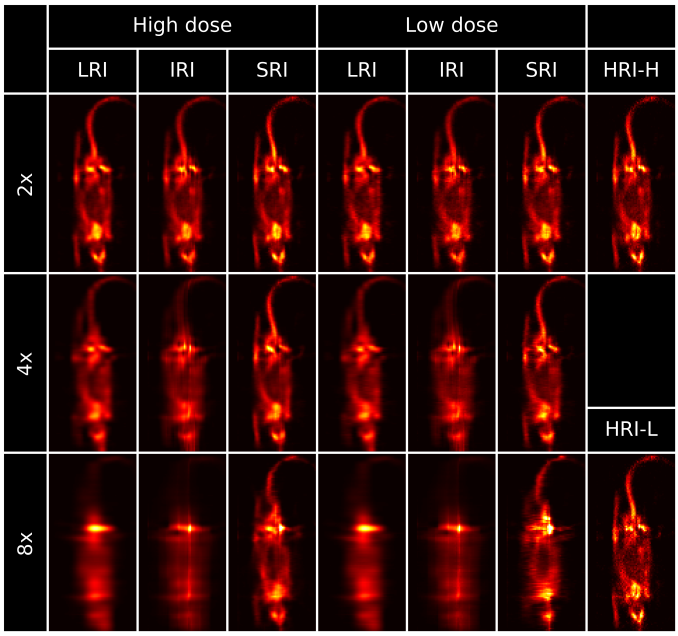

Fig. 16. Reconstructed images of mice data. HRI-H and HRI-L are HRIs for high dose and low dose respectively.

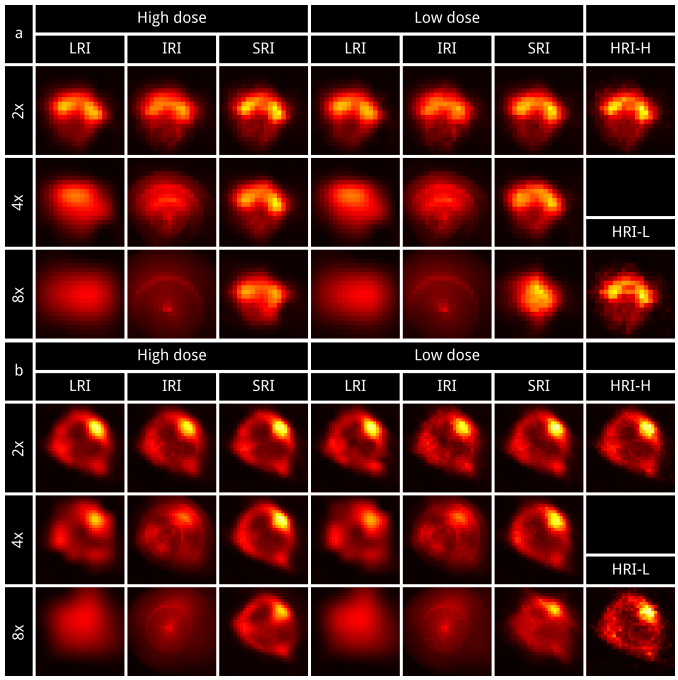

Fig. 17. Zoomed transverse images of brain (a) and heart(b) for reconstructed mice data. HRI-H and HRI-L are HRIs for high dose and low dose respectively.

Zero padding, which is widely used in CNNs, results in high line shape artifacts and poor performance of the DRSSRN near the patch edges. Unlike in the natural image, boundary pixels in sinograms may influence the whole reconstruction image. Thus we cropped boundary pixels to increase overall performance.

We did not use batch normalization (BN) introduced by [46] that is widely used in CNNs for classification and many other tasks. We tested using BN and found that it needed much more time for each training step, and sometimes degraded the performance after convergence, which was also observed in [47].

To further accelerate learning, we used the concatenate exponential linear unit (CELU) as the activation function. It is similar to CReLU introduced by [48] with ELUs replacing the rectifier linear units (ReLUs). ELUs have negative values that allow them to push mean unit activations closer to zero as BN does, but with lower computational complexity. Using concatenate positive and negative activations, we can use fewer variables while obtaining a higher model capacity.

Multiscale structure were used to making it possible to train deep SISR network for large down-sampling ratio. These short-cut paths help gradients propagate to front layers of network, SRB would outputs zeros which made SISR network degenerate to interpolation without multiscale loss.

B. Learning sinograms vs. learning images

Most SISR algorithms were designed for transaxial images in medical image processing tasks. In this work we applied the network on sinograms instead. The comparison shows that SISR in projection domain has better performance.

One of the reasons may stand in the difference processing natural images and medical images. The absolute intensity of the natural image is not as important as that of medical images. People often use random light and contrast to generate more natural images for data augmentation. Besides, the position of organs matters more in medical images. These two distinctions make argumentation difficult on medical images.

On the other hand, the projection domain has some good properties that the natural image does not have. Firstly, we can easily use periodical padding to generate more projections. On the other hand, directly using projection domain data keeps maximum information of a PET scan. Learning reconstructed images cannot recover information lost during the reconstruction process. Furthermore, the most important advantage of CNN over deep neural networks (DNNs) is that the former shares weights from different patches of the image, which makes it much more efficient in learning local features and weaker in learning global features. The difference between a low-resolution sinogram (LRS) and high-resolution sinogram (HRS) is local, but the difference between a low-resolution reconstructed image (LRI) and a high-resolution reconstructed image (HRI) is global.

C. Reconstruction method

Study of influence of different reconstruction methods on our approach is beyond the scope of this paper. However, we have tested Simultaneous Algebraic Reconstruction Technique (SART) reconstruction instead of FBP for all sinograms. Results show similar trends in FBP, but the improvement of SRI is smaller.

We have also tested tuning cut-off frequencies (18 tested cut-offs between 0.05 and 0.5) and filter types (Ram-Lak, Hamming, Hanning, Cosine and Shepp-Logan) for FBP reconstruction. The performance in terms of PSNR and SSIM

changed and in some cases HRI showed better performance than LRI and IRI. However, in all test cases, SRI achieved better performance than HRI, IRI and LRI.

Even though the performance gain is different when using different reconstruction methods, the proposed method can easily be plugged into the existing reconstruction framework because we learned a pure post processing method on sinograms. This may enable the use of simpler reconstruction methods such as FBP instead of iterative methods in some cases to reduce the reconstruction time.

D. Transfer learning

As mentioned in II-C, the DRSSRN suffers from limited training data and poor labeling due to high noise. In order to make the DRSSRN practical in medical imaging applications, we proposed a method of transfer learning to solve these problems. The method used pre-trained network for weight initialization to help train network in a new domain, based on the fact that sinograms share many low-level features independent of the scan geometry and dose level. Our test results validated the transfer learning from analytically simulated sinograms to Monte Carlo simulated sinograms, and then to mice data. This suggested that the DRSSRN might be generalized on different data domains. With carefully designed MC simulations, the difficulty of gathering real scan data might be overcome.

Transfer learning also reduced computational cost for training in new data domain. In our configuration, it costs around 150 GTX 1080 Ti GPU hours to convergence when using noise-free APSs, while transferring it to noisy data and other different types of data costs only 20 to 48 GPU hours.

E. Future works

Even though the DRSSRN works well for different types of data in this work, some details remain to be improved.

When generating training dataset with MC simulations, we did not simulate attenuation or scattering to reduce computational cost. Since there is no big difference between effects caused by attenuation and scattering on large crystals and thin crystals, we ignored them in the current stage. We will try MC simulated data which is more close to realistic data.

And for network architecture, we chose to use 64 filters and 20 residual blocks from grid search with only 4 possible numbers of filters (16, 32, 64 and 128) and 4 possible numbers of residual blocks (5, 10, 20, 40). The architecture described in section II-C was one of the 16 combinations mentioned above. There might be architectures with better performance. The optimal network architecture depends on data and number of samples. We had developed a similar network on roughly 1/4 data used and achieved the best performance with 5 residual blocks. In the future, we may use residual blocks with different configurations, share weights between SRBs, try different down-sampling ratio, and study the effect of number of samples and different types of data to optimize the network. Once optimized, the proposed method will also be validated with clinical data.

VII. Conclusion

In this work, we developed a novel DRSSRN to enhance the image quality for the PET scanning system with coarse pixelated crystals. The proposed DRSSRN is a dedicated deep CNN-based SISR network for sinograms. It borrows SISR methods for natural images. But with periodical padding, cropping and deprecation of BN structure, the performance is improved especially for sinograms. Besides, multi-scale loss short-cuts improve the DRSSRN for large down-sampling ratio. Furthermore using transfer learning makes training for noisy data and small dataset possible. Comparisons suggest that the proposed method is able to achieve comparable image qualities with 4 times larger crystals.

With external training data as prior knowledge of sinograms from thin crystals, the proposed method outputs high-resolution sinograms without requiring other information such as the sequence of images or other modality data, thus can be easily plugged into existing normal reconstruction procedures. This work may provide an alternative solution to the PET system design, with reduced cost, shorter scanning time and improved image quality.

Acknowledgment

This work is supported by the National Natural Science Foundation of China (NSFC) 51627807. The authors would like to thank supports from Center for HPC, Shanghai Jiao Tong University.

References

[1] S. S. Gambhir, J. Czernin, J. Schwimmer, D. H. Silverman, R. E. Coleman, and M. E. Phelps, "A tabulated summary of the fdg pet literature," *Journal of nuclear medicine*, vol. 42, no. 5 suppl, pp. 1S–93S, 2001.

[2] A. Moretti, A. Gorini, and R. Villa, "Affective disorders, antidepressant drugs and brain metabolism," *Molecular psychiatry*, vol. 8, no. 9, p. 773, 2003.

[3] F. M. Bengel, B. Permanetter, M. Ungerer, S. Nekolla, and M. Schwaiger, "Non-invasive estimation of myocardial efficiency using positron emission tomography and carbon-11 acetate–comparison between the normal and failing human heart," *European Journal of Nuclear Medicine and Molecular Imaging*, vol. 27, no. 3, pp. 319–326, 2000.

[4] J. Zhou and J. Qi, "Fast and efficient fully 3d pet image reconstruction using sparse system matrix factorization with gpu acceleration," *Physics in medicine and biology*, vol. 56, no. 20, p. 6739, 2011.

[5] Y. Yang, Y.-C. Tai, S. Siegel, D. F. Newport, B. Bai, Q. Li, R. M. Leahy, and S. R. Cherry, "Optimization and performance evaluation of the micropet ii scanner for in vivo small-animal imaging," *Physics in medicine and biology*, vol. 49, no. 12, p. 2527, 2004.

[6] M. Ito, S. J. Hong, and J. S. Lee, "Positron emission tomography (pet) detectors with depth-of-interaction (doi) capability," *Biomedical Engineering Letters*, vol. 1, no. 2, p. 70, 2011.

[7] P. Bruyndonckx, C. Lemaître, D. Schaart, M. Maas, M. Krieguer, O. Devroede, S. Tavernier *et al.*, "Towards a continuous crystal apd-based pet detector design," *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, vol. 571, no. 1, pp. 182–186, 2007.

[8] W. Yonggang, C. Xinyi, and L. Deng, "Depth of interaction estimation using artificial neural network for continuous crystal pet detector," in *Nuclear Science Symposium and Medical Imaging Conference (NSS/MIC), 2012 IEEE*. IEEE, 2012, pp. 1260–1263.

[9] A. González, P. Conde, A. Iborra, A. Aguilar, P. Bellido, R. García-Olcina, L. Hernández, L. Moliner, J. Rigla, M. Rodríguez-Álvarez *et al.*, "Detector block based on arrays of 144 sipms and monolithic scintillators: A performance study," *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, vol. 787, pp. 42–45, 2015.

[10] J. Cabello, P. Barrillon, J. Barrio, M. Bisogni, A. Del Guerra, C. Lacasta, M. Rafecas, H. Saikouk, C. Solaz, P. Solevi *et al.*, "High resolution detectors based on continuous crystals and sipms for small animal pet," *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, vol. 718, pp. 148–150, 2013.

[11] S. Ahn and R. M. Leahy, "Analysis of resolution and noise properties of nonquadratically regularized image reconstruction methods for pet," *IEEE transactions on medical imaging*, vol. 27, no. 3, pp. 413–424, 2008.

[12] S. Tong, A. Alessio, and P. Kinahan, "Noise and signal properties in psf-based fully 3d pet image reconstruction: an experimental evaluation," *Physics in Medicine & Biology*, vol. 55, no. 5, p. 1453, 2010.

[13] A. Rahmim, J. Qi, and V. Sossi, "Resolution modeling in pet imaging: theory, practice, benefits, and pitfalls," *Medical physics*, vol. 40, no. 6Part1, 2013.

[14] D. Wallach, F. Lamare, G. Kontaxakis, and D. Visvikis, "Super-resolution in respiratory synchronized positron emission tomography," *IEEE transactions on medical imaging*, vol. 31, no. 2, pp. 438–448, 2012.

[15] J. A. Kennedy, O. Israel, A. Frenkel, R. Bar-Shalom, and H. Azhari, "Improved image fusion in pet/ct using hybrid image reconstruction and super-resolution," *International journal of biomedical imaging*, vol. 2007, 2007.

[16] M. Bevilacqua, A. Roumy, C. Guillemot, and M. line Alberi Morel, "Low-complexity single-image super-resolution based on nonnegative neighbor embedding," in *Proceedings of the British Machine Vision Conference*. BMVA Press, 2012, pp. 135.1–135.10.

[17] S. Schulter, C. Leistner, and H. Bischof, "Fast and accurate image upscaling with super-resolution forests," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2015, pp. 3791–3799.

[18] J. Yang, Z. Wang, Z. Lin, S. Cohen, and T. Huang, "Coupled dictionary training for image super-resolution," *IEEE transactions on image processing*, vol. 21, no. 8, pp. 3467–3478, 2012.

[19] R. Zeyde, M. Elad, and M. Protter, "On single image scale-up using sparse-representations," in *International conference on curves and surfaces*. Springer, 2010, pp. 711–730.

[20] 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.

[21] R. Fang, T. Chen, and P. C. Sanelli, "Towards robust deconvolution of low-dose perfusion ct: Sparse perfusion deconvolution using online dictionary learning," *Medical image analysis*, vol. 17, no. 4, pp. 417–428, 2013.

[22] X.-Y. Cui, Z.-G. Gui, Q. Zhang, H. Shangguan, and A.-H. Wang, "Learning-based artifact removal via image decomposition for low-dose ct image processing," *IEEE Transactions on Nuclear Science*, vol. 63, no. 3, pp. 1860–1873, 2016.

[23] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," in *Advances in neural information processing systems*, 2012, pp. 1097–1105.

[24] C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke, and A. Rabinovich, "Going deeper with convolutions," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2015, pp. 1–9.

[25] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 770–778.

[26] N. Zhang, J. Donahue, R. Girshick, and T. Darrell, "Part-based r-cnns for fine-grained category detection," in *European conference on computer vision*. Springer, 2014, pp. 834–849.

[27] R. Girshick, "Fast r-cnn," in *Proceedings of the IEEE international conference on computer vision*, 2015, pp. 1440–1448.

[28] S. Ren, K. He, R. Girshick, and J. Sun, "Faster r-cnn: Towards real-time object detection with region proposal networks," in *Advances in neural information processing systems*, 2015, pp. 91–99.

[29] H. Chen, Y. Zhang, W. Zhang, P. Liao, K. Li, J. Zhou, and G. Wang, "Low-dose ct denoising with convolutional neural network," in *Biomedical Imaging (ISBI 2017), 2017 IEEE 14th International Symposium on*. IEEE, 2017, pp. 143–146.

[30] W. Shen, M. Zhou, F. Yang, C. Yang, and J. Tian, "Multi-scale convolutional neural networks for lung nodule classification," in *International Conference on Information Processing in Medical Imaging*. Springer, 2015, pp. 588–599.

[31] A. Teramoto, H. Fujita, O. Yamamuro, and T. Tamaki, "Automated detection of pulmonary nodules in pet/ct images: Ensemble false-positive reduction using a convolutional neural network technique," *Medical physics*, vol. 43, no. 6, pp. 2821–2827, 2016.

[32] J. Dolz, N. Reyns, N. Betrouni, D. Kharroubi, M. Quidet, L. Massoptier, and M. Vermandel, "A deep learning classification scheme based on augmented-enhanced features to segment organs at risk on the optic region in brain cancer patients," *arXiv preprint arXiv:1703.10480*, 2017.

[33] C. Dong, C. C. Loy, K. He, and X. Tang, "Image super-resolution using deep convolutional networks," *IEEE transactions on pattern analysis and machine intelligence*, vol. 38, no. 2, pp. 295–307, 2016.

[34] J. Kim, J. Kwon Lee, and K. Mu Lee, "Accurate image super-resolution using very deep convolutional networks," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2016, pp. 1646–1654.

[35] X. Hong, Y. Zan, F. Weng, Z. Zhao, and Q. Huang, "Enhancing pet image quality via single image super resolution with deep residual learning," in *the proceedings of The 14th International Meeting on Fully Three-Dimensional Image Reconstruction in Radiology and Nuclear Medicine*, 2017, pp. 673–676.

[36] B. Xu, C. Liu, Y. Dong, R. Tang, Y. Liu, H. Yang, M. Chen, C. Li, Z. Shen, Y. Dong *et al.*, "Performance evaluation of a high-resolution tof clinical pet/ct," *Journal of Nuclear Medicine*, vol. 57, no. supplement 2, pp. 202–202, 2016.

[37] D.-A. Clevert, T. Unterthiner, and S. Hochreiter, "Fast and accurate deep network learning by exponential linear units (elus)," *arXiv preprint arXiv:1511.07289*, 2015.

[38] M. Abadi, A. Agarwal, P. Barham, E. Brevdo, Z. Chen, C. Citro, G. S. Corrado, A. Davis, J. Dean, M. Devin *et al.*, "Tensorflow: Large-scale machine learning on heterogeneous distributed systems," *arXiv preprint arXiv:1603.04467*, 2016.

[39] D. Kingma and J. Ba, "Adam: A method for stochastic optimization," *arXiv preprint arXiv:1412.6980*, 2014.

[40] W. Segars, G. Sturgeon, S. Mendonca, J. Grimes, and B. M. Tsui, "4d xcat phantom for multimodality imaging research," *Medical physics*, vol. 37, no. 9, pp. 4902–4915, 2010.

[41] Z. Wang, E. P. Simoncelli, and A. C. Bovik, "Multiscale structural similarity for image quality assessment," in *Signals, Systems and Computers, 2004. Conference Record of the Thirty-Seventh Asilomar Conference on*, vol. 2. IEEE, 2003, pp. 1398–1402.

[42] H. C. Gifford, M. A. King, D. J. de Vries, and E. J. Soares, "Channelized hotelling and human observer correlation for lesion detection in hepatic spect imaging," *The Journal of Nuclear Medicine*, vol. 41, no. 3, p. 514, 2000.

[43] S. Jan, G. Santin, D. Strul, S. Staelens, K. Assie, D. Autret, S. Avner, R. Barbier, M. Bardies, P. Bloomfield *et al.*, "Gate: a simulation toolkit for pet and spect," *Physics in medicine and biology*, vol. 49, no. 19, p. 4543, 2004.

[44] S. Healthcare, "Biograph mct technology brochure," 2008.

[45] M. Defrise, P. E. Kinahan, D. W. Townsend, C. Michel, M. Sibomana, and D. F. Newport, "Exact and approximate rebinning algorithms for 3-d pet data," *IEEE transactions on medical imaging*, vol. 16, no. 2, pp. 145–158, 1997.

[46] S. Ioffe and C. Szegedy, "Batch normalization: Accelerating deep network training by reducing internal covariate shift," in *International Conference on Machine Learning*, 2015, pp. 448–456.

[47] B. Lim, S. Son, H. Kim, S. Nah, and K. M. Lee, "Enhanced deep residual networks for single image super-resolution," *arXiv preprint arXiv:1707.02921*, 2017.

[48] W. Shang, K. Sohn, D. Almeida, and H. Lee, "Understanding and improving convolutional neural networks via concatenated rectified linear units," in *International Conference on Machine Learning*, 2016, pp. 2217–2225.