Super-resolution reconstruction for parallel-beam SPECT based on deep learning and transfer learning: a preliminary simulation study

Zhibiao Cheng[1]^, Junhai Wen[1], Jun Zhang[1], Jianhua Yan[2]

[1]Department of Biomedical Engineering, School of Life Science, Beijing Institute of Technology, Beijing, China; [2]Shanghai Key Laboratory of Molecular Imaging, Jiading Central Hospital, Shanghai University of Medicine and Health Sciences, Shanghai, China

Contributions: (I) Conception and design: Z Cheng, J Wen, J Yan; (II) Administrative support: J Wen; (III) Provision of study materials or patients: Z Cheng, J Wen, J Yan; (IV) Collection and assembly of data: Z Cheng, J Zhang; (V) Data analysis and interpretation: Z Cheng, J Yan; (VI) Manuscript writing: All authors; (VII) Final approval of manuscript: All authors.

Correspondence to: Junhai Wen. Department of Biomedical Engineering, School of Life Science, Beijing Institute of Technology, Beijing 100081, China. Email: wenjh@bit.edu.cn; Jianhua Yan. Shanghai Key Laboratory of Molecular Imaging, Jiading Central Hospital, Shanghai University of Medicine and Health Sciences, Shanghai 201318, China. Email: Jianhua.yan@gmail.com.

Background: Single-photon emission computed tomography (SPECT) is widely used in the early diagnosis of major diseases such as cardiovascular disease and cancer. High-resolution (HR) imaging requires HR projection data, which typically comes with high costs. This study aimed to obtain HR SPECT images based on a deep learning algorithm using low-resolution (LR) detectors.

Methods: A super-resolution (SR) reconstruction network based on deep learning and transfer learning for parallel-beam SPECT was proposed. LR SPECT sinograms were converted into HR sinograms. Training data were designed and generated, including digital phantoms (128×128 pixels), HR sinograms (128×128 pixels), and LR sinograms (128×64 pixels). A series of random phantoms was first used for pre-training, and then the extended cardiac-torso (XCAT) phantom was used to fine-tune the parameters. The effectiveness of the method was evaluated using an unknown cardiac phantom. To simulate a wide range of noise levels, the total count levels of the projection were normalized to 1e7 (100%), 1e6 (10%), and 1e5 (5%). Finally, the training sets for different count levels were generated. Transfer learning was employed to accelerate the training.

Results: The proposed network was validated on the simulation data sets using different Poisson noise levels. The quantitative values of the peak signal-to-noise ratio (PSNR) and structural similarity index measure (SSIM) indicators of the reconstructed images were improved compared to those recorded using the benchmark methods. Using the proposed method, an image resolution comparable to that of images reconstructed from the HR projection could be achieved.

Conclusions: Based on deep learning and transfer learning, an SR reconstruction network in the projection domain of the parallel-beam SPECT was developed. The simulation results under a wide range of noise levels evidenced the potential of the proposed network to improve SPECT resolution for LR detector scanners.

Keywords: Parallel-beam single-photon emission computed tomography (SPECT); super-resolution (SR); deep learning; transfer learning; projection

^ ORCID: 0000-0001-7636-5311.

Page 2 of 17

Cheng et al. SR reconstruction for parallel-beam SPECT

Introduction

Single-photon emission computed tomography (SPECT) is commonly used in the early diagnosis and treatment of diseases such as cancer and cardiovascular and cerebrovascular diseases (1). To reconstruct high-resolution (HR) SPECT images, the corresponding HR projection data need to be obtained. HR detectors need to be placed in the SPECT system, which leads to high hardware costs and may hinder the wide application of SPECT. Compared with HR detectors, low-resolution (LR) detectors have the advantages of low cost and ease of popularization; however, they also have the disadvantage of low imaging resolution. Therefore, methods for obtaining HR SPECT images with LR projection are worthy of research.

The reconstruction of SPECT images from projection data is an inverse problem. Traditional SPECT reconstruction methods mainly include analytical filtered back-projection (FBP) algorithms and iterative reconstruction algorithms, such as the algebraic reconstruction technique (ART), maximum likelihood expectation maximization (MLEM), ordered-subset expectation maximization (OSEM), and maximum a posteriori (MAP) (2). Although FBP algorithms have the advantage of being fast, their use requires a choice to be made between HR and noise. In comparison, reconstructed images obtained using iterative algorithms are more accurate, because they benefit from more complex system modeling; however, they entail much higher computational costs. The application of regularization terms can optimize the imaging quality, but there is still no unified standard for parameter selection and the regularization of terms. The premise of applying these methods is to obtain the HR projection data.

Researchers have taken various approaches to studying super-resolution (SR) in SPECT. One of the most important strategies is to apply multi-image SR technology on the image domain (3). In the projection domain, an HR projection can be calculated from a series of LR projections containing different information, and then the image is reconstructed using the reconstruction method (4-6). However, the design of these methods is based on a perfect, error-free calibration system, and, in practice, errors in motion and registration may introduce uncertainty. The recovery of HR images from LR images is also an ill-posed problem (7). Another strategy to obtain HR images is to use single-image SR technology, such as an SR method based on sparse representation and local linear regression (8-10). These methods are applied based on the assumption that the HR image manifold and the LR image manifold

have similar local geometric structures. The LR image block close to the test image block in the training set is searched, and the corresponding HR image block is used to reconstruct the test image. The effectiveness of these methods is limited by the ability of the underlying features to express the details of the HR image.

In recent years, researchers have applied deep learning technology to nuclear medical image reconstruction (2,11). Some learning reconstruction networks have been proposed to realize the direct mapping between projection data and activity images (12-16). The reconstruction step has been represented as two fully connected layers or a deep convolutional encoder-decoder network. Deep learning has often been used to optimize traditional reconstruction methods. For example, the U-Net network has been used to improve the quality of FBP and MLEM reconstruction images (17,18). Some researchers have integrated artificial intelligence (AI) structures into iterative reconstruction steps (19-23), which is a time-consuming process. In SR reconstruction, because medical images and natural images have different data distributions, the direct use of deep neural networks trained on natural image datasets for medical image processing may not result in good-quality images. Medical images are usually characterized by similar structures. For example, in the projection domain, all sinograms can be regarded as being composed of sinogram lines under different parameters. Furthermore, when SR technology is applied to medical images, it is necessary to improve the image resolution or to denoise and ensure the correctness of the reconstructed images, especially in the case of lesions. Hong *et al.* (24) proposed a residual convolutional neural network (CNN) to predict HR positron emission tomography (PET) sinogram data from LR sinogram data, thereby improving the effectiveness of learning local feature information. In the image domain, the convolutional encoder-decoder network was proven to be effective in image detail restoration and achieved end-to-end mapping from LR PET images to HR PET images (25,26). Kim *et al.* (27) generated HR SPECT images from LR SPECT images using a dense, block-based CNN. Furthermore, magnetic resonance-guided PET image reconstruction technology (mainly post-processing) can provide more prior details to improve the resolution of anatomical images (28); however, multi-modality images with good registration are not easy to obtain.

The purpose of this study was to propose a learning approach based on residual learning and transfer learning to improve the imaging resolution of SPECT scans using the
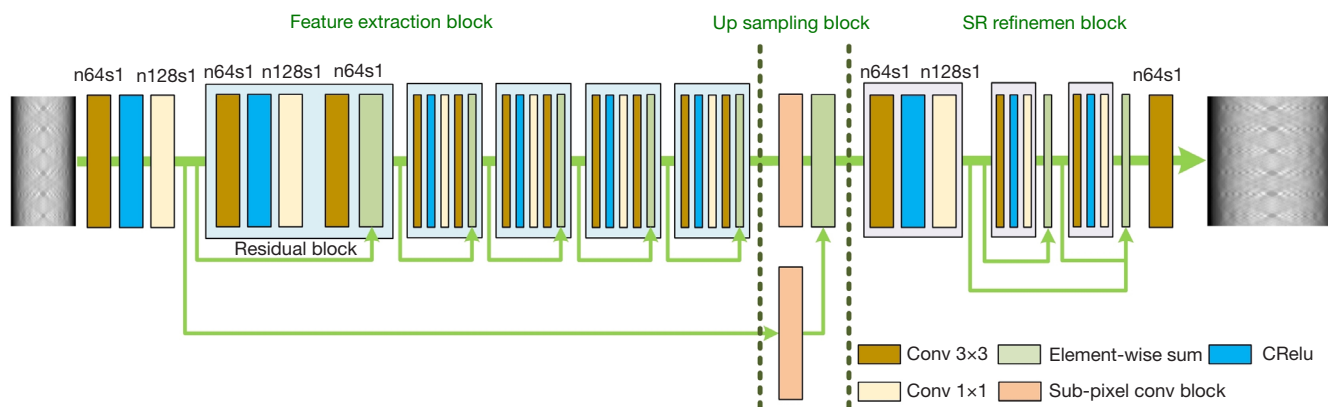
Figure 1 The structure of the proposed network. Sub-pixel conv block, upsampling operation. SR, super-resolution; n64s1, number of channels 64, stride 1; CRelu, concatenated rectified linear unit.

LR detector. We focused on two-dimensional (2D) parallel-beam SPECT reconstruction. The input of the proposed network was an LR sinogram (128×64 pixels) and the output was an SR sinogram (128×128 pixels). The benefit of this design was that it converted the 2D relationship between the LR sinogram and the HR sinogram into a one-dimensional relationship, further simplifying the complexity of network learning. The proposed network was verified on the phantom datasets under different Poisson noise levels. To mimic clinical situations, the Monte Carlo method was used to obtain sinograms of an unknown cardiac defect to verify the generalizability of the proposed method. Once the mapping between the LR sinogram and the HR sinogram was established, the SR sinogram could be obtained to improve the resolution of the final reconstruction. We present the following article in accordance with the Transparent Reporting of a multivariable prediction model for Individual Prognosis Or Diagnosis (TRIPOD) reporting checklist (available at https://atm.amegroups.com/article/view/10.21037/atm-21-4363/rc).

Methods

It is difficult to effectively train an SR network to map the relationship between a total (HR) projection domain and the image domain. Compared with HR projection, the nonlinear relationship between an LR projection image and a reconstructed image is more complex and requires the fit of a deeper, multi-layered, convolutional network. Limited by computer hardware and a large number of parameters, direct training may face the difficulties of a slow convergence speed and a difficult network convergence.

Therefore, the problem was transformed into mapping the relationship between the training LR projection and the HR projection. After the SR projection image had been obtained, the SR SPECT image could be reconstructed using the conventional reconstruction algorithm.

SR network for projection

As explained above, it is difficult to apply an SR network trained with natural images to projection data, due to different pattern distributions. In the first part of our study, the goal was to train an SR network, G, to estimate the corresponding 2D HR sinogram for a given 2D LR input sinogram, described as:

$$\hat{\theta}_G = \arg\min_{\theta_G} \frac{1}{N} \sum_{n=1}^{N} loss\left(G_{\theta_G}\left(I_n^{LR}\right), I_n^{HR}\right) \quad [1]$$

where $\theta_G = \{W_{1:L}; b_{1:L}\}$ denotes all parameters (weights and biases of all L-layer network) of the network, *loss* denotes the loss function, and I_n^{LR}, I_n^{HR}, and $G_{\theta_G}\left(I_n^{LR}\right)$ denote the 2D LR sinogram, the 2D HR sinogram, and the 2D SR sinogram, respectively. The single-channel, gray-scale sinogram was described using real-valued tensors of $H \times (W/r) \times 1$ and $H \times W \times 1$ in size, respectively, where r represents the scale factor and N represents the number of paired data in the training set, $n=1, \dots, N$.

As shown in *Figure 1*, the proposed network was divided into three parts: a feature extraction block, an upsampling block, and an SR refinement block. In the feature extraction block, after the application of a convolution layer with a 3×3 filter, five identical residual blocks were serially used to extract and aggregate the low-frequency features in the LR sinogram (29). Another residual structure was designed,
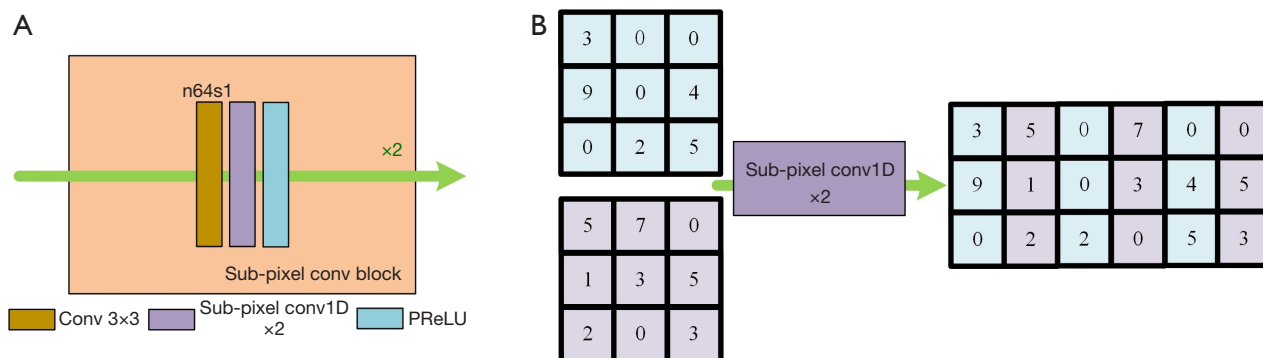
Figure 2 The sub-pixel convolution block. (A) The structure of the one-dimensional sub-pixel convolution block. (B) An example demonstrating the operation of the sub-pixel convolution layer. n64s1, number of channels 64, stride 1; PReLU, parametric rectified linear unit.

and the global features were extracted by setting jumper connections outside the five residual blocks. The function of the upsampling block was to merge the above-mentioned features and convert them to an HR size. The final SR sinogram was obtained by designing SR refinement blocks with dense connectivity, which further refined the high-frequency features.

In the feature extraction block, after the application of a convolution layer with a 3×3 filter, a concatenated rectified linear unit (CReLU), and a convolution layer with a 1×1 filter were designed (30). The residual block consisted of two convolution layers with 3×3 filters, one convolution layer with a 1×1 filter, and a CReLU as the activation function (31,32). With the deepening of the network, the CReLU network could capture positive- and negative-phase information, thus avoiding the redundancy of the convolution kernel. The CReLU is described as:

$$CReLU(x) = Concat\left[ReLU(x),\ ReLU(-x)\right] \qquad [2]$$

$$ReLU(x) = max(0, x) \qquad [3]$$

where, x is the input of the neuron. The rectified linear unit (ReLU, or linear rectification function) is a common activation function in artificial neural networks (33).

The number of feature maps in the feature extraction block was set to 64, and the batch normalization layer was discarded. After CReLU, the number of feature maps was doubled; subsequently, a convolutional layer with a 1×1 filter was used to restore the number of feature maps to 64. In addition to the five residual blocks, the independent jumper connection could better retain the global characteristics and eliminate the redundant information obtained (34).

In the upsampling block, 2 one-dimensional, sub-pixel convolution blocks were used to implement multi-channel parallel upsampling. The inputs of two sub-pixel convolution blocks were the output of the residual blocks and the output of the first convolution layer, respectively (35). The outputs of the two sub-pixel convolution blocks were fused by the element-wise sum operation. As shown in *Figure 2A*, before being input to the sub-pixel convolution layer, the feature maps from the residual block output had to be calculated using the convolution layer with 3×3 filters. The number of channels of the sub-pixel convolution layer was equal to 64×2. After application of the sub-pixel convolution layer, a common parametric rectified linear unit (PReLU) activation function was applied (36,37). A demonstration of how to operate the sub-pixel convolution layer is shown in *Figure 2B*.

Instead of directly taking the output of the sub-pixel convolution block as the result, after upsampling, we input the feature maps into the SR refinement block. The SR refinement block included three convolutional blocks and a convolutional layer. The convolutional block included a convolution layer with a 3×3 filter, a CReLU, and a convolution layer with a 1×1 filter. The setting of the number of feature maps was consistent with the description in the feature extraction block. The dense connection mechanism was applied, and the specific performance requirement was for each layer to accept all the previous layers as its additional input, with a view to achieving feature reuse and improving efficiency (38). The number of output channels of the last convolutional layer with a 3×3 filter was designed to be 1, and there was no activation function.

Digital phantoms and preprocessing

As it is difficult to collect a large number of clinical datasets,
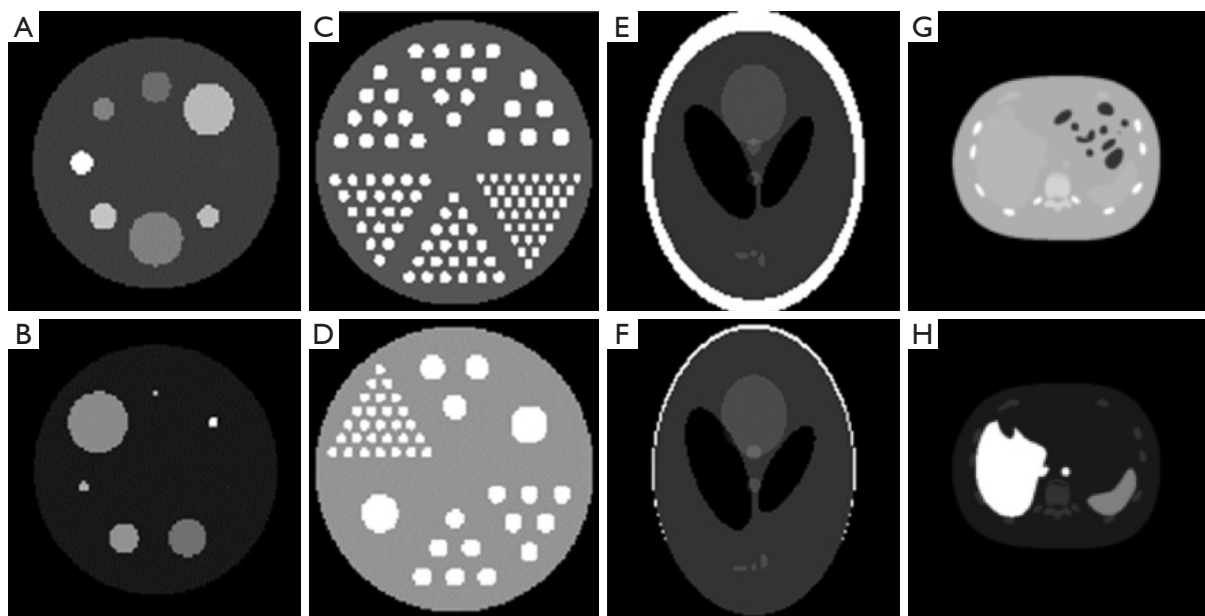
Figure 3 Example display of digital phantoms and XCAT phantoms participating in training. (A,B) Random display of Jaszczak-like phantoms. (C,D) Random display of Derenzo-like phantoms. (E,F) Random display of Shepp Logan-like phantoms. (G) Random display of XCAT attenuation maps. (H) Random display of XCAT exact source image. XCAT, extended cardiac-torso.

this study used digital phantoms to generate training sets. The advantage of using a digital phantom is that it has a unique ground truth image. Through network training, images that are infinitely close to ground truth images can be generated, which cannot be achieved using clinical datasets. A possible disadvantage of using digital phantoms is the difficulty of including all possibilities. However, projections of different medical images have similar structures (which can be regarded as the combination of sinusoids); this allows a large number of digital phantoms to be pre-trained and then further trained on medical data through transfer learning (39). The advantage of transfer learning is that re-learning from scratch can be avoided, and the learning efficiency of the model can be accelerated and optimized.

In this study, 100,000 Shepp Logan-like phantoms, 100,000 Derenzo-like phantoms, and 100,000 Jaszczak-like phantoms were generated. For the Shepp Logan-like phantoms, we used multiple parameter values (position, major axis, minor axis, rotation, and intensity) of different ellipses in the standard Shepp Logan-phantom as the expected values and the expected 1/10 as the variance. The normal distribution was used to randomly generate phantoms in batches. The number of point sources in the six triangular regions in the Derenzo-like phantoms was random and could be selected from the integer

range [1–9]. The intensity values of the background and point sources were random decimals and could be selected from the range (0.0–1.0). Similarly, the number of sources in the Jaszczak-like phantoms was random and could be selected from the integer range [7–12]. The values of radius and intensity of the source were selected in accordance with the mathematical uniform distribution (0.0–1.0) (24). All phantoms were treated as active images, and the size was uniformly set to 128×128 pixels. Noise-free HR sinograms (128×128 pixels) could be obtained from the activity image (128×128 pixels) by applying the forward projection method using the system matrix. To obtain more sampling details, the number of sampling angles was not reduced at this stage. Average pooling with windows 1×2 was used on the HR sinogram to simulate the LR sinogram (128×64 pixels). The number of viewing angles of the sinogram remained unchanged, but the number of bins was reduced by half. To simulate the SPECT scan after the radiotracer injection, the total count level of the projection was normalized to 1e7 (100%). Poisson noise was introduced into the noise-free sinogram, and datasets with Poisson noise were generated. In each category, 90% of the phantoms were used for training and the other 10% were used for testing. *Figure 3A-3F* shows a random selection of different digital phantoms involved in the training process.

Page 6 of 17

Cheng et al. SR reconstruction for parallel-beam SPECT

As shown in *Figure 3G,3H*, to further optimize the generalizability of the network in specific fields, the extended cardiac-torso (XCAT) phantoms were validated (40). In total, 40 sets of XCAT phantoms were applied, and the CT images were converted to corresponding SPECT images based on tissue segmentation (41). In this study, the liver was chosen as the target organ. Due to the limitations of the graphics processing unit (GPU) of memory and network complexity, the image size was reduced to 128×128×160 from 256×256×160 pixels. For the XCAT phantom, the HR sinogram (128×128 pixels) was obtained from the activity image (128×128 pixels) using a forward projection method and the system matrix and attenuation map. The way the LR sinogram was obtained was consistent with the way the LR sinogram of the digital phantom was obtained. We used the sinogram with a total count level of 1e7 (100%), which was regarded as high count data. To simulate a wide range of noise levels, the total count level of the sinogram was normalized to 1e6 (10%) and 1e5 (5%) in turn, and training sets at different count levels were generated. Thirty-five sets of XCAT phantoms were used as the training set; three sets of XCAT phantoms were used as the validation set; and two sets of XCAT phantoms were used as the test set. Finally, an extra-cardiac phantom was simulated by the Monte Carlo method for SPECT imaging (42). The active and the attenuation images of each slice were set to 128×128 pixels, and the image size was set to 320×320 mm^2. The HR detector unit size was set to 2.5 mm^2; the collimating hole diameter was set to 2.5 mm; the hole length was set to 100 mm; and the height of the scintillation crystal was set to 50 mm. To verify the performance of the proposed method at different doses, the number of photons emitted at each angle was set to 2.7×10^6 and 2×10^5 in turn. The performance of the proposed method was verified when the cardiac phantom demonstrated tissue defects.

Training details and parameters

The input of the proposed SR network was 128×64 pixels and the output was 128×128 pixels. The periodicity of the sinogram (with a period of 2pi) was used to enhance the training data and avoid the boundary error caused by convolution. In the vertical axis direction, three identical sinograms were spliced from end to end, and then randomly cropped to obtain a sinogram of the same size as before splicing. Further, random flips and rotations were used to enhance the training data. Before inputting into the network, we used standardization to preprocess the training data, as follows:

$$\left[\hat{I}_n^{LR}, I_n^{HR}\right] = \frac{\left[I_n^{LR}, I_n^{HR}\right] - mean\left(I_n^{LR}\right)}{standard\ deviation\left(I_n^{LR}\right)} \qquad [4]$$

where, $\left[I_n^{LR}, I_n^{HR}\right]$ represents the paired 2D LR sinogram and 2D HR sinogram, and $\left[\hat{I}_n^{LR}, I_n^{HR}\right]$ represents the standardized paired 2D LR sinogram and 2D HR sinogram.

It was difficult to directly obtain the mapping of the noise-added LR and HR sinograms. Using transfer learning, a progressive training program was designed. In Step 1, because extensive clinical datasets could not be obtained during the training phase, the proposed network was trained on the noise-free digital phantom dataset. The network was trained for a total of 30 epochs (2.8e5 iterations). The learning rate was initially set to 0.0001 and decreased to 0.5 times the original rate every 20 epochs. The batch size was set to 32. The Adam optimizer was also used (43), with the settings β_1 =0.9 and β_2 =0.999. The objective of the proposed method was to minimize the difference between the real SR ground truth sinogram and the output sinogram. L1loss was selected as the loss function (44).

In Step 2, the purpose was to provide initial parameters for Step 3 and accelerate the training speed, which was a transitional step. To obtain a better network parameter, in the noise digital data training, the noise-added LR digital sinogram was used as the input and the noise-free HR digital sinogram was used as the label. Transfer learning was used, and the parameters trained with noise-free digital data were used as the initial training parameters of this network. For this step, the network was trained at 30 epochs (2.8e5 iterations). The parameter settings were set to be consistent with Step 1.

In Step 3, the proposed network was trained using the noise-add XCAT phantom dataset. Compared with the digital phantom, the XCAT phantom experiment was closer to the clinical case scenario. To simulate a wide range of noise levels, the total count level of the projection was normalized to 1e7 (100%), 1e6 (10%), and 1e5 (5%). The networks under the different noise levels were trained separately, in order of 1e7, 1e6, and 1e5. Transfer learning was used again, in order to avoid training from scratch, and the parameters trained for the noise-added digital phantom were used as the initial training parameters for the training step (1e7). The training at the other noise levels was deduced by analogy. The network (1e7 counts)

was trained for a total of 130 epochs (2.4e4 iterations); the network (1e6 counts) was trained for a total of 130 epochs (2.4e4 iterations); the network (1e5 counts) was trained for a total of 150 epochs (2.7e4 iterations). Other than the initial learning rate, which was set to 0.00001, the other parameter settings were set to be consistent with Step 1.

Combined with Compute Unified Device Architecture (CUDA) Toolkit v.10.0, and Pytorch v.1.20, Python code was used to implement the proposed method on the NVIDIA GeForce RTX 2080 GPU. Due to the use of the complete convolution neural network, the trained neural network was suitable for input of any size.

Statistical analysis

To quantitatively evaluate the performance of the proposed algorithm, two standard metrics, the peak signal-to-noise ratio (PSNR), and the structural similarity index measure (SSIM), which are widely used to evaluate image quality, were used to compare the prediction image with the ground truth image (45).

The PSNR is widely used to evaluate prediction accuracy, and is an objective criterion for image evaluation. The formula is as follows:

$$PSNR = 10\log_{10}\frac{\max\left(u_n, u_{original_n}\right)^2}{\frac{1}{N}\sum_{n=1}^{N}\left(u_n - u_{original_n}\right)^2} \quad [5]$$

where u_n and $u_{original_n}$ are the pixel values of the reconstructed image and the reference image, respectively, and N is the total number of pixels in the image. A larger PSNR reflects better image quality.

The SSIM is an indicator used to evaluate the similarity between two images, and is calculated as follows:

$$SSIM = \frac{2\bar{u}_{original}\bar{u}\left(\sigma_{u_{original}u} + c_2\right)}{\left(\bar{u}_{original}^2 + \bar{u}^2 + c_1\right)\left(\sigma^2 + \sigma_{original}^2 + c_2\right)} \quad [6]$$

where $\bar{u}_{original}$ and $\sigma_{u_{original}}$ denote the mean value and variance of the reference image, respectively, $\bar{u}$ and σ denote the mean value and variance of the reconstructed image, respectively, $\sigma_{u_{original}u}$ is the covariance of the two images, c_1 and c_2 are constants. The closer the SSIM value is to 1, the more similar are the images.

For the images obtained by different methods under different noise levels, the average PSNR and SSIM were calculated in turn, along with the standard deviation (SD). The average activity concentration in the high-uptake area and low-uptake background area were also calculated, along with the SD (16).

Results

Results of simple digital phantom

Figure 4 shows the sinograms and reconstructed images of the untrained noise-free digital phantom under different methods. By setting the intensity values of adjacent positions on the horizontal axis as equal, the LR sinogram could be changed to the HR size. Using our method, more sinogram details could be restored than could be with the LR sinograms (128×128) or the bicubic interpolation sinograms; the difference between our results and HR sinogram was the smallest. We opted to take a transverse tangent (representing the detector direction, line 54 of sinogram) from all the sinograms, which is depicted in *Figure 5*. The SR sinogram obtained using our method was closest to the HR sinogram; more sinogram details could be restored to obtain better reconstructed image quality than with the LR sinogram or the bicubic sinogram. Using these sinograms, the images were reconstructed by the ART reconstruction algorithm. In the case of noise-free images, compared with the images reconstructed by the LR sinogram and the bicubic sinogram, the image reconstructed by the SR sinogram was clearer and the details were more completely restored.

Figure 6 shows the sinograms and reconstructed images of the untrained noise-added digital phantom under the different methods in Step 2. Since it is difficult to directly train SR networks under noise, the label applied here is noise-free HR sinogram, and the results shown here are transitional. Compared with the LR sinogram and the bicubic sinogram, the SR sinogram had a better signal-to-noise ratio and could be reconstructed with a quality close to the original phantom. Compared with the traditional interpolation method, the performance of our method was closer to that of the noise-added sinogram, and effectively improved the reconstruction quality without introducing obvious errors. To quantitatively evaluate the performance of the reconstructions, we computed the average PSNR and the SSIM (standard is noise-free HR sinogram), as shown in *Table 1*, where a higher value corresponds to a better reconstruction; the results in *Table 1* belong to the held-out test set of the digital phantom. The parameters obtained in this part can be considered as intermediate parameters. We directly applied the network parameters obtained under this data to the noise-add XCAT phantom and used them as the initialization parameters under the new dataset.

Page 8 of 17

Cheng et al. SR reconstruction for parallel-beam SPECT
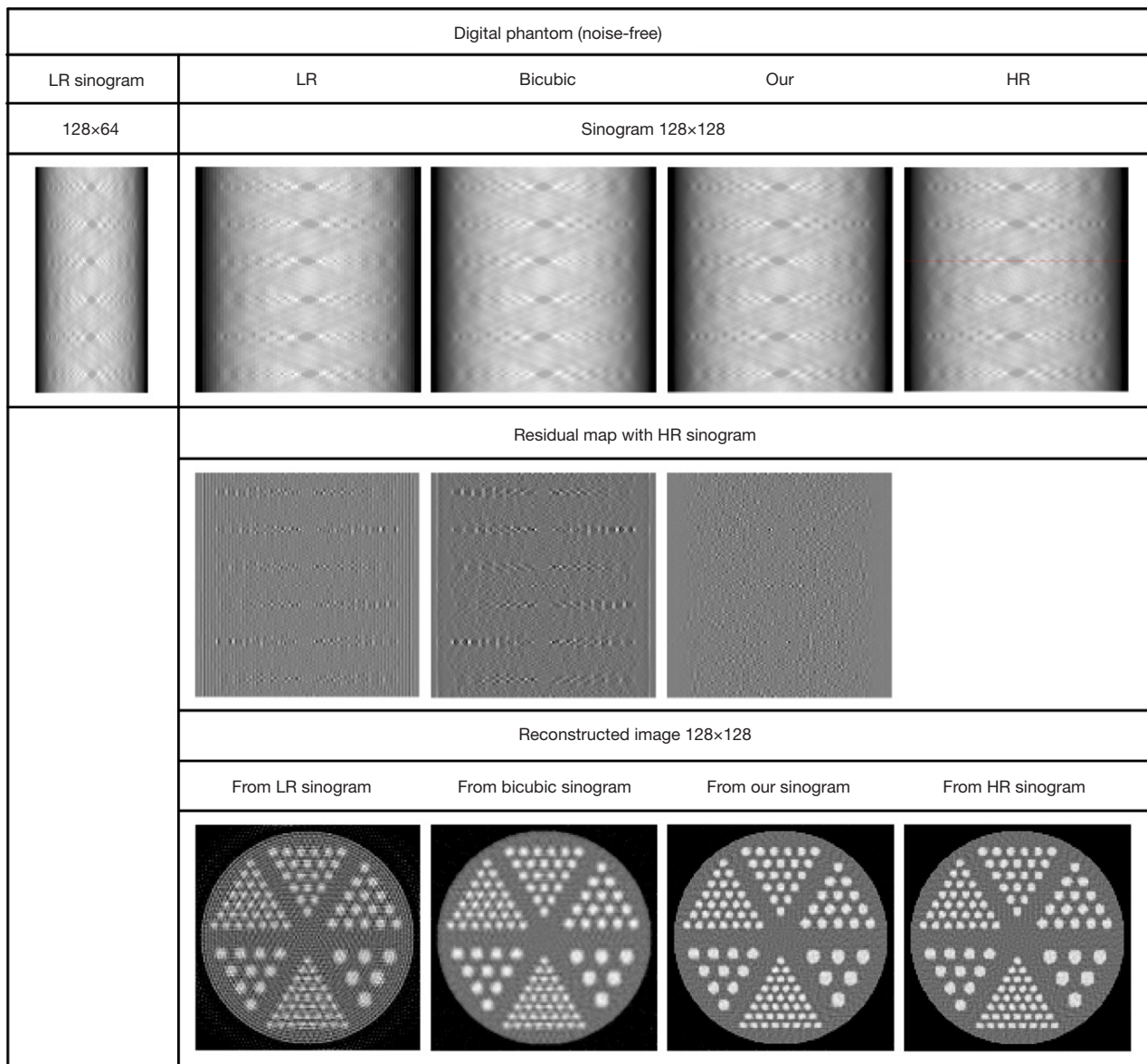
Figure 4 Noise-free sinogram and reconstructed image using different methods. LR, low-resolution; HR, high-resolution.

Results of digital XCAT phantom

Figure 7 shows the results of the liver phantom under different methods. For Step 3, the input was the noise-added LR sinogram, and the label was noise-added HR sinogram. The sinogram was reconstructed using the OSEM reconstruction method. *Table 2* shows the mean and SD of the PSNR and the SSIM of different results in the held-out test set of the XCAT phantom under different noise levels. The results obtained using our proposed method achieved better numerical evaluation and visual effects. For the images presented in *Figure* 7, we also calculated the average activity concentration in the high-uptake liver area and low-uptake background area, along with the SD. The calculated values are presented in *Table 3*. At all noise levels, the average activity concentration values in our images were closer to the exact values than those images for both the high-uptake and low-uptake regions. Compared with the traditional interpolation method, the error between the result obtained by our method and the real HR sinogram was smaller, which was reflected in the more detailed restoration of the reconstructed image.
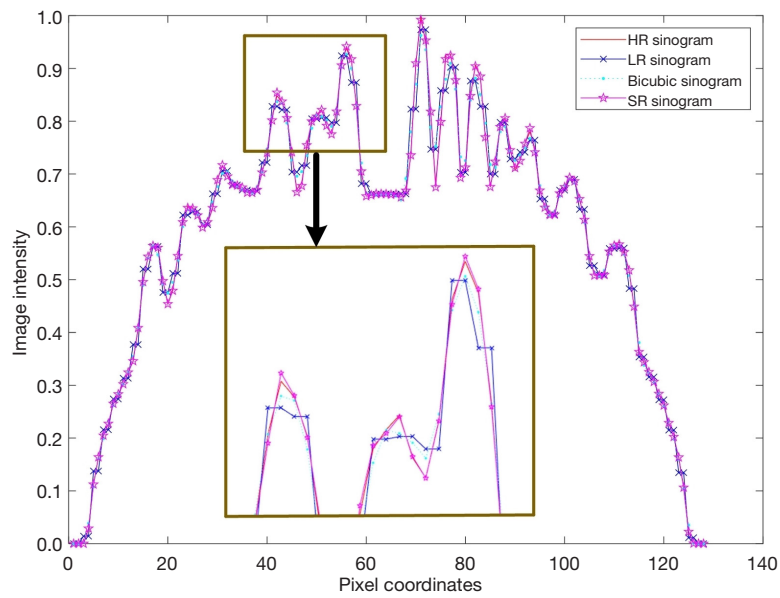
Figure 5 Typical line profile of the projection obtained using different methods. HR, high-resolution; LR, low-resolution; SR, super-resolution.
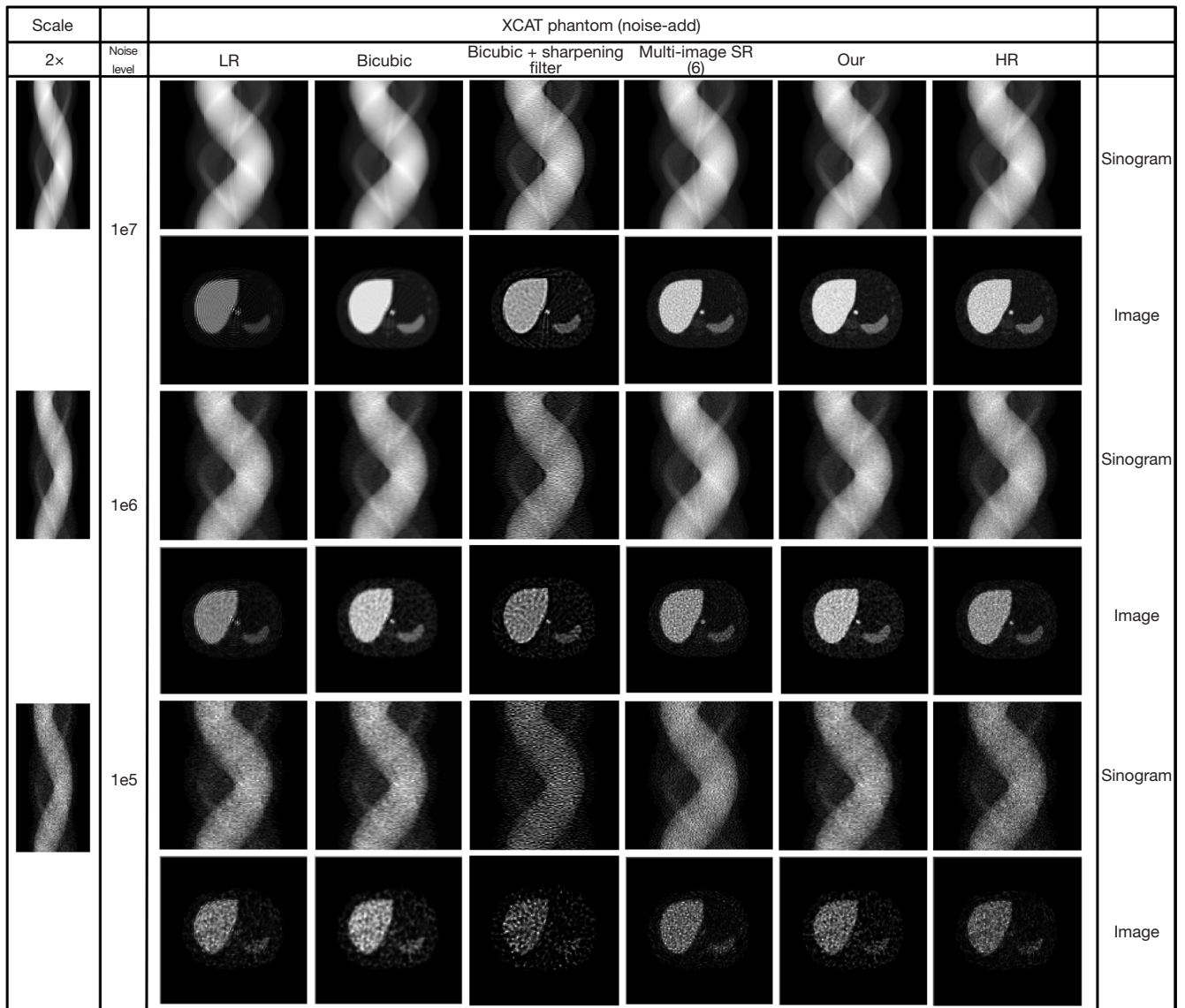

Figure 6 Noise-added sinogram and reconstructed image obtained using different methods. LR, low-resolution; HR, high-resolution.

Table 1 Average PSNR and SSIM of the different results (the held-out test set of digital phantoms)

Image type	Index	Noise-free			Noise-add 1e7 counts		
		LR	Bicubic	Our	LR	Bicubic	Our
Sinogram	PSNR	36.77	41.78	52.59	33.05	35.28	42.01
	SSIM	0.9087	0.9521	0.9944	0.8477	0.8931	0.9776
Image	PSNR	19.19	25.30	34.72	16.41	21.44	26.86
	SSIM	0.5091	0.8081	0.9673	0.4130	0.6054	0.8366

PSNR, peak signal-to-noise ratio; SSIM, structural similarity index measure; LR, low-resolution.

Figure 7 Noise-added sinogram and reconstructed image under different methods for the XCAT phantom (6). XCAT, extended cardiac-torso. LR, low-resolution; SR, super-resolution; HR, high-resolution.

Table 2 Mean and SD (in parenthesis) of PSNR and SSIM for the different liver results (the held-out test set of XCAT phantom)

Noise level	Method	Sinogram		Image	
		PSNR	SSIM	PSNR	SSIM
Noise-added 1e7 counts	LR	35.22 (1.64)	0.9403 (0.0060)	31.97 (3.84)	0.9191 (0.0269)
	Bicubic	38.87 (1.08)	0.9580 (0.0052)	37.30 (3.11)	0.9322 (0.0437)
	Sharpening	32.19 (0.78)	0.8911 (0.0247)	32.45 (2.99)	0.9140 (0.0208)
	Multi-image SR (6)	40.62 (1.14)	0.9702 (0.0143)	38.76 (3.16)	0.9604 (0.0147)
	Our	42.44 (1.00)	0.9802 (0.0051)	42.84 (3.17)	0.9777 (0.0095)
Noise-added 1e6 counts	LR	32.52 (0.75)	0.8958 (0.0177)	31.24 (2.58)	0.8856 (0.0150)
	Bicubic	33.49 (1.03)	0.8997 (0.0222)	33.56 (2.27)	0.9006 (0.0162)
	Sharpening	24.67 (1.01)	0.7494 (0.0396)	29.39 (1.96)	0.8593 (0.0144)
	Multi-image SR (6)	32.05 (1.13)	0.8686 (0.0345)	31.31 (2.43)	0.8899 (0.0201)
	Our	33.95 (1.10)	0.9112 (0.0205)	35.01 (2.37)	0.9170 (0.0143)
Noise-added 1e5 counts	LR	26.53 (1.01)	0.8045 (0.0244)	28.25 (2.67)	0.8332 (0.0274)
	Bicubic	26.37 (1.06)	0.7851 (0.0293)	28.52 (2.74)	0.8195 (0.0335)
	Sharpening	18.64 (1.02)	0.6660 (0.0321)	24.59 (3.19)	0.7811 (0.0267)
	Multi-image SR (6)	25.20 (1.08)	0.7254 (0.0411)	26.75 (2.79)	0.7871 (0.0370)
	Our	27.01 (1.03)	0.8106 (0.0258)	28.94 (2.64)	0.8429 (0.0278)

SD, standard deviation; PSNR, peak signal-to-noise ratio; SSIM, structural similarity index measure; XCAT, extended cardiac-torso; LR, low-resolution; SR, super-resolution.

Although the sinogram processed by the Sharpening filter {Laplace operator [0 1 0; 1 –4 1; 0 1 0] was used} obtained more edge information, there was no obvious improvement in the reconstruction result. The multi-image SR method described in (6) could obtain the effect close to the real HR sinogram, but it would require more LR detectors, making it more expensive. The ring artifacts caused by LR acquisition in the reconstructed image could be more effectively eliminated by our SR method. Compared with the images reconstructed by the LR sinogram and the bicubic sinogram, our results also showed reduced noise. It should be noted that under higher noise, the performance of the network decreased, but it still obtained a better boundary in the reconstructed image.
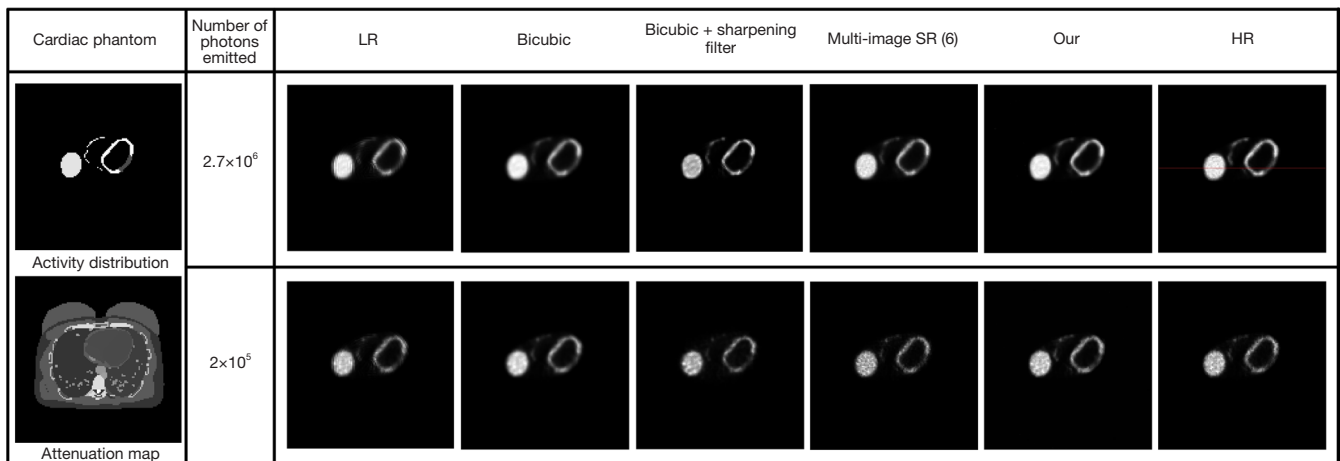
Results of the cardiac phantom

Figure 8 shows a slice of the cardiac phantom under different methods. In this part of our study, we used the new phantom and the new projection method to test our network. A perfusion defect was simulated on the anterolateral myocardium of the left ventricle, and the extent of the defect was one-third of the normal activity. When the number of photons emitted at each angle was set to 2.7×10^6, we used the corresponding network parameters under the noise level of 1e7. When the number of photons emitted at each angle was set to 2×10^5, we used the corresponding network parameters under the 1e5 noise level. The experimental results show that the reconstructed image obtained by our method was closer to the HR sinogram reconstructed image. Compared with the method in (6), our reconstruction results had less noise; specifically, the reconstructed image from the LR sinogram had obvious ring artifacts, the image after bicubic processing was more blurred, and the result after sharpening processing was obviously non-uniform. For the first row in *Figure 8*, a transverse tangent (line 64 of image) was taken from all the reconstructed images, as shown in *Figure 9*. We have zoomed in on two details, which show that the proposed method was closer to the HR result and demonstrated a better boundary. *Table 4* shows the evaluation index scores of the cardiac phantom in *Figure 8* under different dose levels. *Table 5* shows the numerical comparison between the different methods depicted in *Figure 8* in the uptake

Page 12 of 17

Cheng et al. SR reconstruction for parallel-beam SPECT

Table 3 Numerical comparison between the different methods, as depicted in *Figure* 7, used in the uptake regions, by activity mean value and SD

Noise level	Method	High-uptake region mean activity (SD)	Low-uptake region mean activity (SD)
Noise-added 1e7 counts	True	1	0.1
	LR	0.9955 (0.1872)	0.1041 (0.0585)
	Bicubic	0.9966 (0.0291)	0.1028 (0.0202)
	Sharpening	1.0366 (0.1275)	0.0922 (0.0572)
	Multi-image SR (6)	1.0004 (0.0361)	0.1030 (0.0156)
	HR	1.0016 (0.0909)	0.1007 (0.0266)
	Our	1.0010 (0.0510)	0.1006 (0.0239)
Noise-added 1e6 counts	True	1	0.1
	LR	0.9891 (0.1664)	0.1043 (0.0492)
	Bicubic	0.9905 (0.0926)	0.1023 (0.0399)
	Sharpening	1.0122 (0.2795)	0.0926 (0.0885)
	Multi-image SR (6)	0.9954 (0.1864)	0.1020 (0.0505)
	HR	0.9956 (0.1766)	0.1022 (0.0461)
	Our	0.9953 (0.1422)	0.1011 (0.0531)
Noise-added 1e5 counts	True	1	0.1
	LR	0.9916 (0.2997)	0.1098 (0.0865)
	Bicubic	0.9936 (0.2403)	0.1080 (0.0790)
	Sharpening	0.9939 (0.6605)	0.1252 (0.2195)
	Multi-image SR (6)	0.9901 (0.2519)	0.1116 (0.0622)
	HR	0.9981 (0.3845)	0.1101 (0.0961)
	Our	0.9988 (0.3243)	0.1087 (0.0974)

The high-uptake region refers to the liver region and the low-uptake region refers to the background region. SD, standard deviation; LR, low-resolution; SR, super-resolution; HR, high-resolution.

Figure 8 Comparison of cardiac images under different sampling photons (6). LR, low-resolution; SR, super-resolution; HR, high-resolution.
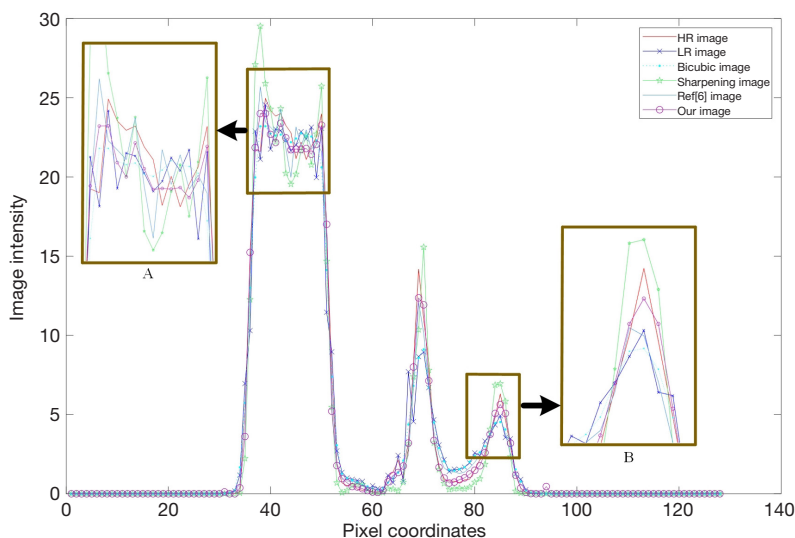
Figure 9 Typical line profile of the projection obtained by different methods for the cardiac phantom (6). HR, high-resolution; LR, low-resolution.

Table 4 Comparison of evaluation indicators (PSNR and SSIM) for different cardiac results

No. of photons emitted at each angle	Method	PSNR	SSIM
$2.7×10^6$	LR	33.98	0.9762
	Bicubic	35.14	0.9797
	Sharpening	32.37	0.9833
	Multi-image SR (6)	38.18	0.9879
	Our	40.61	0.9954
$2×10^5$	LR	36.42	0.9874
	Bicubic	37.08	0.9883
	Sharpening	32.37	0.9814
	Multi-image SR (6)	34.60	0.9842
	Our	37.81	0.9990

PSNR, peak signal-to-noise ratio; SSIM, structural similarity index measure; LR, low-resolution; SR, super-resolution.

regions, by activity mean value and SD. Compared with the HR results, the average activity concentration values obtained using our method were found to be more accurate than those obtained using other methods, and the errors were also smaller than those occurring by other methods. Furthermore, the reconstructed image obtained by our method could better reconstruct the defects in the cardiac tissue.

Discussion

In this study, we have developed a single-image SR method on sinograms for LR SPECT systems. Residual learning and transfer learning were applied and verified on different phantoms. Compared with traditional methods, the projected image and the reconstructed image obtained using our SR method achieved better PSNR and SSIM performances. The proposed SR network performed well for the noisy sinograms. We found that it was almost impossible to completely recover the HR projection under noise (especially at the high-noise level), but our method showed the advantage of reducing the error observed with the HR sinogram. On the basis of preserving the details of the HR image as much as possible, a certain denoising effect was achieved, which meant that the reconstructed image had a better signal-to-noise ratio. We used transfer learning to train the SR network in a noisy environment, step by step. The application of transfer learning accelerated and optimized the convergence speed of the model on the target phantom to a certain extent, and effectively reduced the unknown impact of the model's dependence on data distribution. *Figure 10*, for instance, shows the different performances when applying or not applying transfer learning in the network training in Step 3, comparing the average SSIM on the XCAT dataset (1e7 counts), which

Page 14 of 17

Cheng et al. SR reconstruction for parallel-beam SPECT

Table 5 Numerical comparison between the different methods depicted in *Figure 8* in the uptake regions, by activity mean value and SD

No. of photons emitted at each angle	Method	High-uptake region mean activity (SD)	Low-uptake region mean activity (SD)
$2.7×10^6$	LR	20.30 (4.21)	5.27 (1.54)
	Bicubic	20.39 (3.93)	5.29 (1.54)
	Sharpening	22.30 (4.72)	5.96 (2.38)
	Multi-image SR (6)	20.74 (3.81)	5.46 (1.63)
	HR	20.97 (3.67)	6.01 (1.89)
	Our	20.96 (3.23)	5.97 (1.75)
$2×10^5$	LR	21.44 (4.78)	5.60 (1.73)
	Bicubic	21.59 (4.42)	5.64 (1.70)
	Sharpening	20.35 (6.35)	5.51 (2.17)
	Multi-image SR (6)	21.11 (5.02)	5.48 (2.44)
	HR	21.29 (4.79)	5.63 (2.14)
	Our	21.30 (4.15)	5.88 (1.86)

The high-uptake region refers to the liver region and the low-uptake region refers to the perfusion-defect region of the left ventricle. SD, standard deviation; LR, low-resolution; SR, super-resolution; HR, high-resolution.
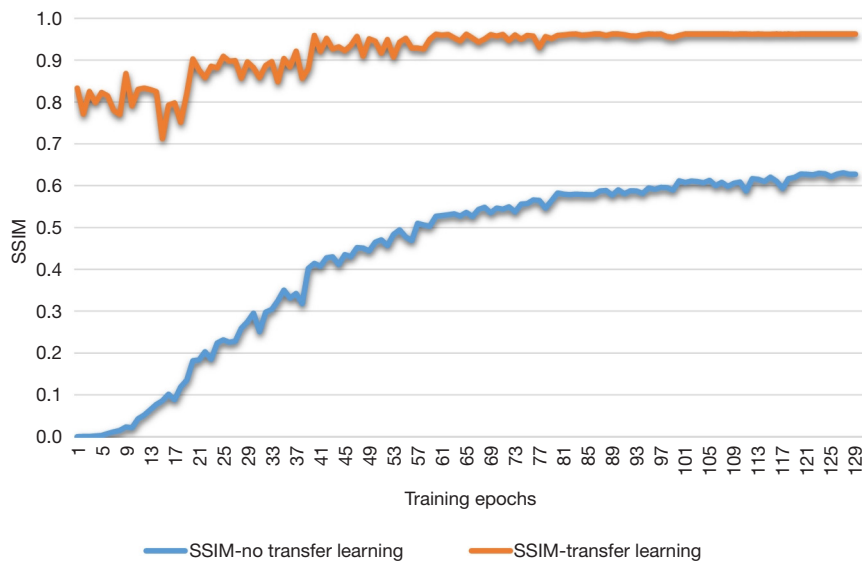
Figure 10 Comparison of the effect of applying or not applying transfer learning on testing the average SSIM on the XCAT data set (1e7 counts) not involved in training. SSIM, structural similarity index measure; XCAT, extended cardiac-torso.

was not included in the training. It was difficult to obtain an ideal performance by training directly on the data under noise. We applied the trained network to simulate the cardiac imaging of myocardial perfusion defect. Numerical examples showed that the proposed network could reconstruct high-quality images, even if the noise on the sinogram was unknown.

There are some limitations to this study. The first is that the final prediction effort was relatively poor at the positions with large pixel differences in the projection. In particular, under the condition of high noise, it was difficult to obtain the same result as the HR sinogram, and

the error could only be reduced as much as possible. The possible reason for this is that there were very few outliers in the projection data that were too large or too small, which made it difficult to effectively obtain the positional relationship for the data-driven method. However, we found that, compared to the LR reconstruction results, the impact of this error on the reconstruction quality was acceptable. Also, we used simulation methods to obtain a series of data. Limited by the huge simulation calculation time, we did not use the Monte Carlo software to generate training data. However, to simulate different doses in real cases, we verified the performance of the network under different noise levels. Without introducing new errors, we showed that our method can improve resolution and suppress noise.

In practical applications, the detected targets often have similar structures, which has benefits for data-driven deep learning methods and avoids introducing large amounts of invalid data. Furthermore, paired HR sinograms and LR sinograms are almost impossible to obtain. For a known LR SPECT system, the LR sinogram of the target can be obtained by direct acquisition. The corresponding HR sinogram is unknown, but could be obtained by the corresponding HR detector acquisition or Monte Carlo algorithm. Moreover, even if paired data are obtained, it would be difficult to train the network directly under noise. Therefore, in the early stage of network training, the training steps of digital phantoms are indispensable. Since the noise in SPECT can be modeled as Poisson noise, in this paper, the proposed method was verified under different noise levels, which showed that the method has potential application value. On a real device, we would only need to apply transfer learning to fine-tune a small amount of data; then, the true mapping relationship between the LR sinogram and the corresponding HR sinogram could be obtained, thereby improving the reconstruction resolution, which is beneficial for reducing equipment costs.

Conclusions

This study has explored the possibility of using SR technology in a low-cost LR SPECT projection to improve the reconstruction resolution. We first used random phantoms for pre-training, and then used XCAT phantoms for fine-tuning of parameters. The proposed method was verified under different noise levels. Furthermore, an unknown cardiac phantom was used to evaluate the generalizability of the method. Images similar to those

reconstructed from the HR projection using ART and OSEM methods were obtained. The quantitative values of the PSNR and SSIM indicators for the reconstructed image were improved compared to those obtained using the benchmark methods. Our experimental results show that our proposed SR framework can improve imaging resolution and reduce noise. However, the results of this paper are preliminary, and the performance of the proposed method needs to be further verified with clinical data.

Acknowledgments

Footnote

Reporting Checklist: The authors have completed the TRIPOD reporting checklist. Available at https://atm.amegroups.com/article/view/10.21037/atm-21-4363/rc

Data Sharing Statement: Available at https://atm.amegroups.com/article/view/10.21037/atm-21-4363/dss

Conflicts of Interest: All authors have completed the ICMJE uniform disclosure form (available at https://atm.amegroups.com/article/view/10.21037/atm-21-4363/coif). The authors have no conflicts of interest to declare.

Ethical Statement: The authors are accountable for all aspects of the work in ensuring that questions related to the accuracy or integrity of any part of the work are appropriately investigated and resolved.

Open Access Statement: This is an Open Access article distributed in accordance with the Creative Commons Attribution-NonCommercial-NoDerivs 4.0 International License (CC BY-NC-ND 4.0), which permits the non-commercial replication and distribution of the article with the strict proviso that no changes or edits are made and the

Page 16 of 17

Cheng et al. SR reconstruction for parallel-beam SPECT

original work is properly cited (including links to both the formal publication through the relevant DOI and the license). See: https://creativecommons.org/licenses/by-nc-nd/4.0/.

References

1. Zanzonico P. Principles of nuclear medicine imaging: planar, SPECT, PET, multi-modality, and autoradiography systems. Radiat Res 2012;177:349-64.

2. Cheng Z, Wen J, Huang G, et al. Applications of artificial intelligence in nuclear medicine image generation. Quant Imaging Med Surg 2021;11:2792-822.

3. Caramelo FJ, Almeida G, Mendes L, et al. Study of an iterative super-resolution algorithm and its feasibility in high-resolution animal imaging with low-resolution SPECT cameras. In: 2007 IEEE Nuclear Science Symposium Conference Record. Honolulu, HI, USA: IEEE, 2007.

4. Villena JL, Lage E, de Carlos A, Tapias G, Sisniega A, Vaquero JJ, Desco M. A super-resolution feasibility study in small-animal SPECT imaging. In: 2008 IEEE Nuclear Science Symposium Conference Record. Dresden: IEEE, 2008.

5. Massari R, D'Elia A, Soluri A, et al. Super Spatial Resolution (SSR) method for small animal SPECT imaging: A Monte Carlo study. Nucl Instrum Methods Phys Res A 2020;982:164584.

6. Yan Z, Lu Y, Wen J, et al. Super resolution SPECT reconstruction with non-uniform attenuation. Comput Biol Med 2012;42:651-6.

7. Tian J, Ma KK. A survey on super-resolution imaging. Signal Image Video Process 2011;5:329-42.

8. Yang J, Wang Z, Lin Z, et al. Coupled dictionary training for image super-resolution. IEEE Trans Image Process 2012;21:3467-78.

9. Timofte R, De V, Gool LV. Anchored Neighborhood Regression for Fast Example-Based Super-Resolution. In: 2013 IEEE International Conference on Computer Vision (ICCV). Sydney: IEEE, 2013.

10. Timofte R, De Smet V, Gool LV. A+: Adjusted Anchored Neighborhood Regression for Fast Super-Resolution. In: Computer Vision -- Asian Conference on Computer Vision (ACCV) 2014. Singapore: Springer, 2015.

11. Shao W, Rowe SP, Du Y. Artificial intelligence in single photon emission computed tomography (SPECT) imaging: a narrative review. Ann Transl Med 2021;9:820.

12. Zhu B, Liu JZ, Cauley SF, et al. Image reconstruction by domain-transform manifold learning. Nature 2018;555:487-92.

13. Häggström I, Schmidtlein CR, Campanella G, et al. DeepPET: A deep encoder-decoder network for directly solving the PET image reconstruction inverse problem. Med Image Anal 2019;54:253-62.

14. Shao W, Du Y. SPECT image reconstruction by deep learning using a two-step training method. J Nucl Med 2019;60:1353.

15. Shao W, Pomper MG, Du Y. A Learned Reconstruction Network for SPECT Imaging. IEEE Trans Radiat Plasma Med Sci 2021;5:26-34.

16. Shao W, Rowe SP, Du Y. SPECTnet: a deep learning neural network for SPECT image reconstruction. Ann Transl Med 2021;9:819.

17. Kim K, Wu D, Gong K, et al. Penalized PET Reconstruction Using Deep Learning Prior and Local Linear Fitting. IEEE Trans Med Imaging 2018;37:1478-87.

18. Dietze MMA, Branderhorst W, Kunnen B, et al. Accelerated SPECT image reconstruction with FBP and an image enhancement convolutional neural network. EJNMMI Phys 2019;6:14.

19. Kuang Gong, Jiahui Guan, Kyungsang Kim, et al. Iterative PET Image Reconstruction Using Convolutional Neural Network Representation. IEEE Trans Med Imaging 2019;38:675-85.

20. Gong K, Wu D, Kim K, et al. EMnet: an unrolled deep neural network for PET image reconstruction. In: Medical Imaging 2019: Physics of Medical Imaging. San Diego, CA, USA: SPIE, 2019.

21. Lim H, Chun IY, Dewaraja YK, et al. Improved Low-Count Quantitative PET Reconstruction With an Iterative Neural Network. IEEE Trans Med Imaging 2020;39:3512-22.

22. Gong K, Wu D, Kim K, et al. MAPEM-Net: an unrolled neural network for Fully 3D PET image reconstruction. In: 15th International Meeting on Fully Three-Dimensional Image Reconstruction in Radiology and Nuclear Medicine. Philadelphia, PA, USA: SPIE, 2019.

23. Mehranian A, Reader AJ. Model-Based Deep Learning PET Image Reconstruction Using Forward-Backward Splitting Expectation-Maximization. IEEE Trans Radiat Plasma Med Sci 2020;5:54-64.

24. Hong X, Zan Y, Weng F, et al. Enhancing the Image Quality via Transferred Deep Residual Learning of Coarse PET Sinograms. IEEE Trans Med Imaging 2018;37:2322-32.

25. Shiri I, Leung K, Geramifar P, et al. PSFNET: ultrafast generation of PSF-modelled-like PET images using deep

convolutional neural network. J Nucl Med 2019;60:1369.

26. Shiri I, Leung K, Ghafarian P, et al. HiResPET: high resolution PET image generation using deep convolution encoder decoder network. J Nucl Med 2019;60:1368.

27. Kim K, Lee Y. Improvement of signal and noise performance using single image super-resolution based on deep learning in single photon-emission computed tomography imaging system. Nucl Eng Technol 2021;53:2341-7.

28. Schramm G, Rigie D, Vahle T, et al. Approximating anatomically-guided PET reconstruction in image space using a convolutional neural network. Neuroimage 2021;224:117399.

29. He K, Zhang X, Ren S, et al. Deep Residual Learning for Image Recognition. In: 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Las Vegas, NV, USA: IEEE, 2016.

30. Shang W, Sohn K, Almeida D, et al. Understanding and improving convolutional neural networks via concatenated rectified linear units. In: Proceedings of the 33rd International Conference on International Conference on Machine Learning. New York, NY, USA: JMLR, 2016.

31. Zhang K, Zuo W, Chen Y, et al. Beyond a Gaussian Denoiser: Residual Learning of Deep CNN for Image Denoising. IEEE Trans Image Process 2017;26:3142-55.

32. Song C, Yang Y, Wernick M, et al. Approximate 4D Reconstruction of Cardiac-Gated Spect Images Using a Residual Convolutional Neural Network. In: 2019 IEEE International Conference on Image Processing (ICIP). Taipei: IEEE, 2019.

33. Glorot X, Bordes A, Bengio Y. Deep Sparse Rectifier Neural Networks. In: Proceedings of the Fourteenth International Conference on Artificial Intelligence and Statistics. Fort Lauderdale, FL, USA: PMLR, 2011.

34. Tian C, Zhuge R, Wu Z, et al. Lightweight image super-resolution with enhanced CNN. Knowledge-Based Systems 2020;205:106235.

35. Shi W, Caballero J, Huszár F, et al. Real-Time Single Image and Video Super-Resolution Using an Efficient Sub-Pixel Convolutional Neural Network. In: 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Las Vegas, NV, USA: IEEE, 2016.

36. He K, Zhang X, Ren S, et al. Delving Deep into Rectifiers: Surpassing Human-Level Performance on ImageNet Classification. In: 2015 IEEE International Conference on Computer Vision (ICCV). Santiago: IEEE, 2015.

37. Ledig C, Theis L, Huszár F, et al. Photo-Realistic Single Image Super-Resolution Using a Generative Adversarial Network. In: 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Honolulu, HI, USA: IEEE, 2017.

38. Huang G, Liu Z, Maaten LVD, et al. Densely Connected Convolutional Networks. In: 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Honolulu, HI, USA: IEEE, 2017.

39. Zhuang F, Qi Z, Duan K, et al. A comprehensive survey on transfer learning. Proceedings of the IEEE 2021;109:43-76.

40. Segars WP, Sturgeon G, Mendonca S, et al. 4D XCAT phantom for multimodality imaging research. Med Phys 2010;37:4902-15.

41. Khoshakhlagh M, Pirayesh Islamian J, Abedi SM, et al. A Monte Carlo study for optimizing the detector of SPECT imaging using a XCAT human phantom. Nucl Med Rev Cent East Eur 2017;20:10-4.

42. Shi ST, Wen JH, Wang TT, et al. Monte Carlo Simulation of SPECT Projection with Non-uniform Attenuation. In: International Conference on Energy, Environment and Materials Engineering (EEME 2014): Shenzhen. Lancaster, PA, USA: DEStech, 2014.

43. Kingma D, Ba J. Adam: A Method for Stochastic Optimization. In: International Conference on Learning Representations (ICLR). San Diego, CA, USA: Ithaca, 2015.

44. Wang Z, Liu D, Yang J, et al. Deep Networks for Image Super-Resolution with Sparse Prior. In: 2015 IEEE International Conference on Computer Vision (ICCV). Santiago: IEEE, 2015.

45. Hu Z, Wang Y, Zhang X, et al. Super-resolution of PET image based on dictionary learning and random forests. Nucl Instrum Methods Phys Res A 2019;927:320-9.