



Research article

LU-Net: combining LSTM and U-Net for sinogram synthesis in sparse-view SPECT reconstruction

Si Li, Wenquan Ye and Fenghuan Li*

School of Computer Science and Technology, Guangdong University of Technology, Guangzhou 510006, China

* **Correspondence:** Email: fhli20180910@gdut.edu.cn.

Abstract: Lowering the dose in single-photon emission computed tomography (SPECT) imaging to reduce the radiation damage to patients has become very significant. In SPECT imaging, lower radiation dose can be achieved by reducing the activity of administered radiotracer, which will lead to projection data with either sparse projection views or reduced photon counts per view. Direct reconstruction of sparse-view projection data may lead to severe ray artifacts in the reconstructed image. Many existing works use neural networks to synthesize the projection data of sparse-view to address the issue of ray artifacts. However, these methods rarely consider the sequence feature of projection data along projection view. This work is dedicated to developing a neural network architecture that accounts for the sequence feature of projection data at adjacent view angles. In this study, we propose a network architecture combining Long Short-Term Memory network (LSTM) and U-Net, dubbed LU-Net, to learn the mapping from sparse-view projection data to full-view data. In particular, the LSTM module in the proposed network architecture can learn the sequence feature of projection data at adjacent angles to synthesize the missing views in the sinogram. All projection data used in the numerical experiment are generated by the Monte Carlo simulation software SIMIND. We evenly sample the full-view sinogram and obtain the 1/2-, 1/3- and 1/4-view projection data, respectively, representing three different levels of view sparsity. We explore the performance of the proposed network architecture at the three simulated view levels. Finally, we employ the preconditioned alternating projection algorithm (PAPA) to reconstruct the synthesized projection data. Compared with U-Net and traditional iterative reconstruction method with total variation regularization as well as PAPA solver (TV-PAPA), the proposed network achieves significant improvement in both global and local quality metrics.

Keywords: SPECT reconstruction; sparse-view sinogram; sinogram synthesis; U-Net, LSTM

1. Introduction

Single-photon emission computed tomography (SPECT) has been widely used in clinical diagnosis of cardiovascular diseases, bone scans, lung perfusion imaging and lung ventilation imaging [1]. In a SPECT imaging system, a certain concentration of radioactive tracer is administered to the patient. SPECT might provide estimate of the spatial distribution of radioactive tracer inside the patient's body through tomographic reconstruction from the detected emission events. The detector revolving around the human body records the number of single events due to gamma photons emitted by SPECT radioactive tracer distributed inside the body. The gamma photons are detected only if they travel along directions well defined by a collimator. The collection of such emission events can be processed into a set of projection data that has sequence feature along the direction of projection angle. The estimated radiotracer distribution, whose representation under appropriate basis is widely known as the reconstructed image, can provide clinical diagnosis information. However, SPECT radioactive tracer can cause specific radiation damage to the human body and even increase the risk of radiation-induced cancer. Hence, there is a real need to reduce the radiation dose in SPECT studies. Lower radiation dose can be achieved by lowering the activity of administered radiotracer. Under these circumstances, the reconstructed image suffers from various degradations, such as high Poisson noise, severe artifacts, and low spatial resolution, thereby affecting the accuracy of clinical diagnosis [2]. As a result, how to reduce the radiation dose in SPECT imaging under the premise of maintaining the quality of the reconstructed image has become a crucial clinical problem. Model-based iterative reconstruction (MBIR) is one of the major handcrafted methods for reconstructing low-dose projection data [3–5]. The MBIR methods have fine mathematical explanation, and make good use of the projective geometry and physical principle during the data acquisition process. However, the traditional iterative reconstruction methods with certain model parameters, such as the transform filters, the nonlinear shrinkage operator, and the regularization parameters, which need to be determined empirically via prior knowledge, may not adapt to projection data of various phantoms or with various dosage. For instance, the optimal regularization function determined for some body region or normal-dose projection data may not apply to another region or low-dose study. In other words, the traditional iterative methods lack automated means for the optimization of regularization in image reconstruction models.

Over the past several years, the development of deep learning technology has greatly promoted the progress of medical image reconstruction research. In addition to the traditional analytic and iterative image reconstruction methods [6], the combination of handcrafted and learning-based methods has become a promising and attractive trend. In learning-based methodology, the model parameters of either traditional handcrafted framework or neural network architecture are often optimized on given data sets via parallel computing. In the presence of large, balanced image data sets with high-quality labels, the optimized parameters determine robust handcrafted/deep models that can advance the state-of-the-art of many image reconstruction tasks. At present, deep learning techniques can be applied to medical image reconstruction in the following three categories.

(1) Post-processing methods using the reconstructed images as network input to improve image quality. For instances, in the study of low-dose SPECT myocardial perfusion imaging, Ramon et al. [7]

adopted a supervised learning approach to train a convolutional neural network (CNN) with the aim of suppressing imaging noise and thus improving diagnostic accuracy of low-dose SPECT projection data. In reference [8], Chen et al. combined the autoencoder, deconvolution network, and shortcut connection technique into the residual encoder-decoder CNN for low-dose CT image restoration. Based on the encoder-decoder architecture, Zhang et al. [9] further applied the cells of DenseNet to formulate the encoder module, where the reusing of features can increase the depth of a neural network while enhancing its expression ability. The encoder-decoder architecture is capable of stepwise compressing the input image into a feature representation, and then stepwise rebuilding the representation into a full dataset. This architecture is flexible for developing deep models that are effective in noise suppression, streaking artifact removal and structure preservation in low-dose CT study. The image denoising for low-dose CT has also been investigated with a generative adversarial network (GAN) with Wasserstein distance and perceptual similarity [10]. As for MR image processing, Basty et al. [11] proposed a super resolution network based on the U-Net architecture and long short-term memory layers to exploit the temporal aspect of the dynamic cardiac cine MRI data with the aim of recovering a high-resolution cine MRI sequence from low-resolution long-axis images.

(2) End-to-end reconstruction methods generating high-quality images directly from the input projection data or k -space data. In references [12,13], Yang et al. proposed a novel deep network architecture defined over a data flow graph. This architecture was derived from the iterative procedures in Alternating Direction Method of Multipliers (ADMM) for optimizing a compressed sensing (CS)-based MRI reconstruction model, and thus was known as ADMM-Net. In the training phase, all parameters of ADMM-Net, e.g., transform filters and regularization parameters originating from the MRI reconstruction model, and shrinkage operators originating from the ADMM iterative scheme, were discriminatively learned end-to-end using training pairs of under-sampled k -space data and reconstructed image of fully-sampled data. In the testing phase, ADMM-Net had similar computational overhead as ADMM algorithm, but adopted optimized model and algorithm parameters learned from the training data set for CS-based reconstruction task. ADMM-Net was one of the earliest works employing algorithm unrolling in medical image reconstruction. Algorithm unrolling has good potential in medical imaging where large training data sets are difficult to collect and thus conventional deep networks are intractable to train. As another work in this direction, Adler et al. [14] unrolled the primal-dual hybrid gradient algorithm into a deep architecture (PD-Net) for low-dose CT reconstruction task. In particular, the PD-Net substituted primal and dual proximity operators with conventional neural networks, e.g., CNNs, and jointly trained both network and algorithm parameters in an end-to-end fashion. With the similar regard, Zhang et al. [15] unrolled the joint spatial-radon domain reconstruction (JSR) algorithm and approximated the involved inverse operators and proximity operators by CNNs with the same architecture but different set of trainable parameters. The JSR-Net can simultaneously improve the data consistency in both image and Radon domains, which may lead to better image quality for CT reconstruction from incomplete data. Besides algorithm unrolling, there are also literatures discussing end-to-end reconstruction with conventional neural networks. In particular, Häggström et al. [16] presented an end-to-end PET image reconstruction technique (DeepPET) based on a deep encoder-decoder network, which took PET sinogram data as input and directly output high-quality reconstructed images. The DeepPET approach required the use of large and diverse training data sets since this type of model-free approach can result in generalization errors that heavily depend on data sets and do not have a well-defined bound.

(3) Pre-processing methods utilizing pairs of low-dose and normal-dose projection data both in

the sinogram domain for network training. In low-tube current and sparse-view CT imaging, the measured projection data suffers from high-level noise and data deficiency, respectively. Pre-processing approaches such as sinogram denoising and interpolation, followed by advanced iterative reconstruction algorithms, can alleviate the severity of image noise and artifacts to some extent. There are preliminary studies exploring deep neural networks for sinogram pre-processing [17–19,24]. These works mainly combined the residual learning with U-Net architecture to reduce the imaging noise in the low-count sinogram or to synthesize the missing data in the sparse-view sinogram. Later, Shiri et al. [20] followed the similar philosophy and performed PET sinogram interpolation using the convolutional encoder-decoder architecture. In reference [21], Tang et al. proposed a novel sinogram super-resolution generative adversarial network (GAN) model to generate high-resolution sinograms from the low-resolution counterparts. In later references [22,23], GAN was further applied to CT sinogram interpolation. Since the raw projection data is difficult to collect, most existing deep learning-based low-dose studies still focus on the post-processing category. The deep learning-based post-processing methods generally establish a hierarchical architecture composed of many layers and are capable of learning mappings from low-quality reconstructed images to the high-quality counterparts (i.e., labels). This category of reconstruction methods requires sufficient training set of reconstructed images, and thus is often used in conjunction with the low-computational overhead analytic reconstruction algorithms (such as the filtered back-projection algorithm). Under these circumstances, the probabilistic models of the noise and the image usually incorporated in the iterative reconstruction algorithms cannot be fully exploited. The deep learning-based pre-processing methods, on the contrary, can address the above issue, since the image reconstruction is only performed on testing data sets after the time-consuming network training process. The pre-processing category generally learns mappings from low-dose projection data to the normal-dose counterpart. However, the existing deep learning-based pre-processing reconstruction methods seldom exploit the sequence features of sinogram data at adjacent projection view angles.

In this study, we remark that the sequence feature of projection data at adjacent view angles in a sinogram is crucial for synthesis of sparse-view SPECT projection data. With the aim of retaining the sequence continuity of sinograms, we shall propose a novel neural network architecture based on the combination of U-Net and Long Short-Term Memory network (LSTM), dubbed LU-Net. The LSTM modules are added to the down-sampling layers of U-Net to capture the sequence features of the projection data before it is input to the convolution operations. The proposed LU-Net not only exploits the projection data at neighboring view angles, but also makes moderate use of the data at distant view angles, which is more in line with the global property of sinograms and thus may better synthesize the missing data in the sinogram of sparse-view tomographic imaging, as compared to the conventional U-Net. The neural network's learning ability and data fitting ability are fully exploited to determine a mapping from sparse-view projection data to the full-view counterpart. The missing data in the sinogram collected under the sparse-view SPECT imaging is synthesized using the proposed neural network, and then is reconstructed by the preconditioned alternating projection algorithm (PAPA) previously proposed in [4], leading to reconstructed images of superior quality to the iterative reconstruction approach and the conventional neural network-based pre-processing method.

2. Materials and methods

2.1. Problem description

In clinical application, the radiation dose of SPECT imaging can be reduced by lowering the activity of administered radiotracer. In a fixed period of SPECT scanning time, lower radioactivity will inevitably lead to detected count rate reduction on the detector. Low count rates in the sinogram domain are usually accompanied by measured projection data with high Poisson noise, which in turn yield radioactivity distribution estimation with high noise and low contrast if directly reconstructed by conventional image reconstruction approaches. Such low-quality images might not meet the diagnostic criteria. On the other hand, the issue of low count rates may be addressed by reducing the number of projection views during SPECT scanning. Indeed, more imaging time at one projection view leads to higher count rates of this view angle, and thus to projection data with lower Poisson noise. However, we remark that the total SPECT imaging time should not be further lengthened in clinical practice since more total imaging time still brings extra disadvantages, including worse image quality due to more patient motion, more patient discomfort, and less patient throughput. As a result, in a fixed period of SPECT scanning, more imaging time at each projection view can only be achieved by setting fewer projection view angles.

The sparse-view setting of projection view angles in SPECT imaging will lead to incomplete sinogram, and thus the resulting reconstructed images are usually prone to streak artifacts that heavily depend on the amount of sparsity. In order to synthesize the missing view data in the sinogram domain, we develop a U-Net and LSTM-based neural network. The resulting synthetically full-view data can then be reconstructed by the recently proposed PAPA. The combination of neural-network-enabled pre-processing method and PAPA can take advantage of both sinogram synthesis and image prior, thereby providing reconstructed image of low noise, few artifacts and high accuracy.

2.2. Proposed neural network

2.2.1. LU network structure

A deep neural network combining LSTM and U-Net is proposed to synthesize sparse-view sinogram. U-Net is a classical network with encoding and decoding structures. Down-sampling comprises multiple continuous convolution layers and pooling layers, while Up-sampling comprises multiple continuous convolution layers and deconvolution layers. There is concatenation between corresponding layers. It is an effective network structure for medical image processing and is particularly used in medical image segmentation [25].

Conventional U-Net has a problem in sparse-view sinogram synthesis. It does not pay special attention to the sequence feature of projection data which is sequential along the projection view in the sinogram. In order to better utilize the sequence feature of projection data at adjacent view angles, a network architecture combining LSTM and U-Net is proposed. The proposed neural network architecture is shown in Figure 1. The LU network architecture consists of 7 convolution modules and 4 LSTM modules. Each convolution module is composed of two consecutive convolution layers, and an extra convolution layer with one channel is added to the last convolution module. In order to keep the size of input image consistent with that of output image in training process, a zero-padding scheme is used in the convolution process. At the same time, the filter size of each convolution layer is 3×3 , followed by batch normalization layers. The activation function is ReLU. The maximum pooling is used to reduce the size of input image in down-sampling, and the deconvolution is used to expand the image size in up-sampling. Residual learning is applied to accelerate convergence and

reduce streak artifacts in medical images [26,27].

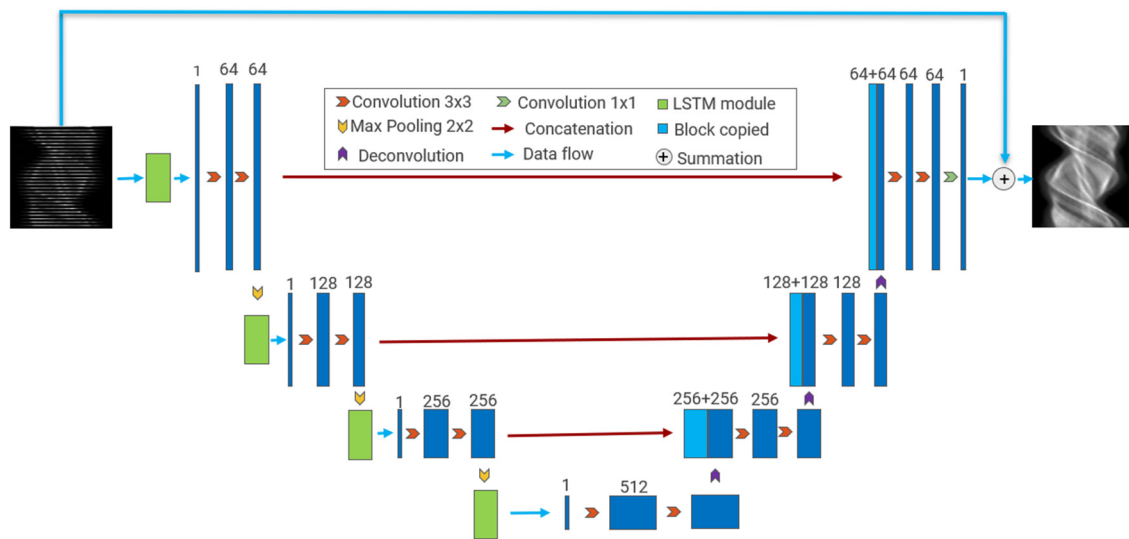


Figure 1. The architecture of LU-Net.

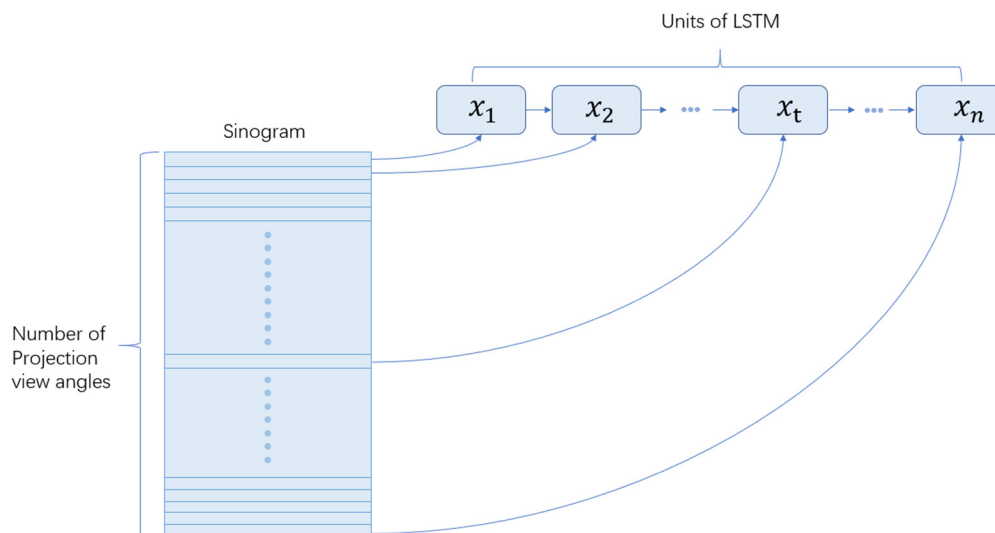


Figure 2. The data flow of sinogram in each LSTM module.

As shown in Figure 1, LSTM modules are added in down-sampling, which can effectively process sequence data. Projection data at each projection view angle in the sinogram can be regarded as time series data and be input into LSTM module for recurrent processing. The detail of sinogram data flow in LSTM module is shown in Figure 2, where x_t is the projection data at projection view angle t . The output of the last recurrent unit is regarded as the output of LSTM module and as the input of convolution layer. In this way, sequence feature of projection data at adjacent view angles can be included in convolution layer, and sparse-view sinogram can be better synthesized.

2.2.2. LSTM module for projection data

LSTM [28] is a special kind of recurrent neural networks (RNN), which is appropriate for processing time series data and learning long-term dependencies. A sinogram is a 2D representation of tomographic scan, where each row represents a single row in the detector array arranged in increasing angular order and thus the whole sinogram is a sequence with respect to projection angle. Moreover, each image pixel generates a sine wave in the sinogram. Therefore, the gray value of each detector element has strong correlation along specific sine curve, instead of with neighbored elements. The sequence feature and sinusoid dependency of sinogram make it a proper candidate for LSTM. The structure of each LSTM unit is shown in Figure 3, which takes x_t , h_{t-1} and C_{t-1} as inputs and is composed of several different functional modules. The details of each module processing the projection data are shown as follows. The variable f_t represents the output of forgetting gate at projection view angle t , which is defined in Eq (1):

$$f_t = \sigma \left(W_f \begin{bmatrix} h_{t-1} \\ x_t \end{bmatrix} + b_f \right). \quad (1)$$

Here x_t is the projection data at projection view angle t as the input to forgetting gate, h_{t-1} is the hidden state at projection view angle $t - 1$, W_f and b_f are the weight and bias of forgetting gate, respectively, and the activation function σ of forgetting gate is a nonlinear mapping sigmoid function. The variable i_t represents the output of input gate at projection view angle t , which is defined in Eq (2):

$$i_t = \sigma \left(W_i \begin{bmatrix} h_{t-1} \\ x_t \end{bmatrix} + b_i \right), \quad (2)$$

$$g_t = \tanh \left(W_g \begin{bmatrix} h_{t-1} \\ x_t \end{bmatrix} + b_g \right). \quad (3)$$

Here the activation function σ is sigmoid function, W_i is the weight of input gate, b_i is the bias of input gate, g_t is a new candidate value vector created in the input gate, which will be added to the cell state later, W_g is the weight of the candidate, b_g is the bias of the candidate. The variable C_t represents the cell state at projection view angle t , which is defined in Eq (4):

$$C_t = f_t * C_{t-1} + i_t * g_t. \quad (4)$$

Here the forgetting gate controls long-term memory, ensuring that the previous projection data saves the information we need, and the input gate controls the current memory, ensuring that irrelevant projection data is not allowed in the cell state. The variable o_t represents the output of output gate at projection view angle t , which is defined in Eq (5).

$$o_t = \sigma \left(W_o \begin{bmatrix} h_{t-1} \\ x_t \end{bmatrix} + b_o \right), \quad (5)$$

$$h_t = o_t * \tanh(C_t). \quad (6)$$

Here W_o and b_o are weight and bias of the output gate, respectively, while h_t is the hidden state of x_t . Note that C_t and h_t are outputs of the underlying LSTM unit.

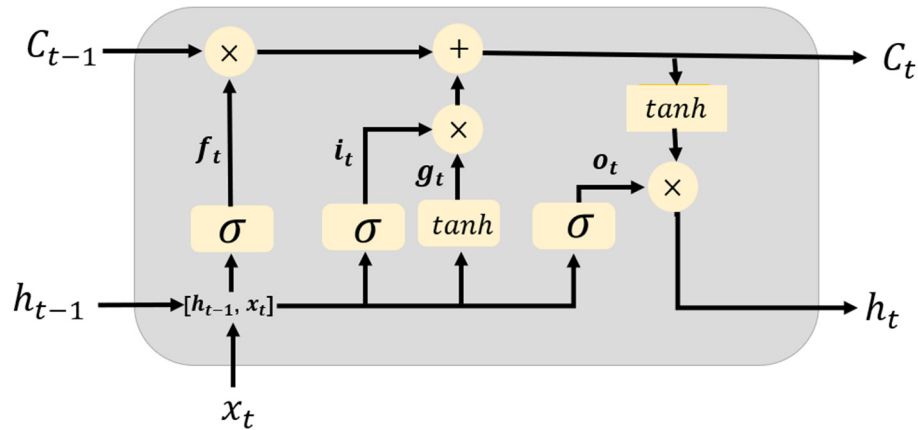


Figure 3. LSTM unit structure.

3. Simulation studies

3.1. Reconstruction methods

In single-photon emission computed tomography (SPECT) system, if the detector dead time can be neglected and no corrections are applied to the raw data, the gamma photon detection at a detector element is a random process that follows the temporal Poisson distribution. In SPECT data acquisition process, the photon detections at different detector elements are independent random events described by Bernoulli process. Hence, the projection data can be modeled as a set of independent Poisson random variables. Indeed, the projection data g detected at m detector elements, which relates to the expected radioactivity distribution representation (i.e., image) f at d pixels in the reconstruction space, can be approximated by the Poisson model [29,30]:

$$g = \text{Poisson}(Af + \gamma). \quad (7)$$

In Eq (7), A is the $m \times d$ SPECT system matrix with the (i, j) th entry equal to the probability of detection of the photon emitted from pixel j of image f by the i th detector element, and γ is the m -dimensional vector of expected counts resulting from the background activity.

Given a realization of the projection data g , the SPECT reconstruction aims to estimate an image f meanwhile suppressing Poisson noise and reducing image artifacts. The reconstruction problem can be formulated via the penalized maximum likelihood criterion, which pursues a regularized estimate by maximizing the sum of log-likelihood function of image f and a negative penalty term [5,10,14]. Since underlying likelihood function is assumed to be Poisson model, the optimization model for SPECT reconstruction usually has the following form [4]:

$$\hat{f} = \underset{f \geq 0}{\text{arg min}} \{ \langle Af, \mathbf{1} \rangle - \langle \ln(Af + \gamma), g \rangle + \lambda \varphi(Bf) \}. \quad (8)$$

In optimization model Eq (8), λ is positive penalty parameter, φ is convex nonnegative function, and B is the regularization operator. Regularization term $\lambda \varphi \circ B$ is introduced to strengthen smoothness of the estimate. Data fidelity function $\langle Af, \mathbf{1} \rangle - \langle \ln(Af + \gamma), g \rangle$, denoted by F below, is the Kullback-Leibler (KL) divergence. Notation $\langle \cdot, \cdot \rangle$ denotes the inner product in Euclidean space,

and $\mathbf{1}$ is a m -dimensional vector with element values of 1.

The iterative algorithm proposed in [4], which is known as PAPA, can be applied to SPECT reconstruction under the penalized maximum likelihood criterion:

$$\begin{cases} \hat{h}^{(k)} := P_+(f^{(k)} - S\nabla F(f^{(k)}) - \mu SB^\top b^{(k)}), \\ b^{(k+1)} = (I - \text{prox}_{(\lambda/\mu)\varphi})\left(b^{(k)} + B\hat{h}^{(k)}\right), \\ f^{(k+1)} = P_+(f^{(k)} - S\nabla F(f^{(k)}) - \mu SB^\top b^{(k+1)}). \end{cases} \quad (9)$$

In above iterative scheme, S is a $d \times d$ diagonal, positive-definite preconditioning matrix that can accelerate convergence of the algorithm, and μ is a positive algorithm parameter. In the case of isotropic total variation regularization, the dual iterate $b^{(k)}$ is defined in the first-order difference transform domain, and $B^\top b^{(k)}$ is regarded as noise in the image domain. The preconditioning matrix S is selected as diagonal matrix $S^{(k)} := \text{diag}(f^{(k)}/A^\top \mathbf{1})$ at the k th iteration of PAPA. The operator P_+ in the iterative scheme is a projection to the first octant. In particular, for any $x \in \mathbb{R}^d$, we have $(P_+(x))_i = \max\{x_i, 0\}$. Finally, motivated by [30], the components of vector $y := \text{prox}_{(\lambda/\mu)\varphi}(z)$ can be computed by the following formula:

$$[y_i, y_{d+i}]^\top = \max\left\{\| [z_i, z_{d+i}]^\top \| - \frac{\lambda}{\mu}, 0\right\} \frac{[z_i, z_{d+i}]^\top}{\| [z_i, z_{d+i}]^\top \|}, \quad (10)$$

$$i = 1, 2, \dots, d.$$

3.2. Data preparation

Experimental data used in this study is obtained by Monte Carlo simulation software SIMIND, which simulates a SIEMENS E.CAM gamma camera with low energy high resolution (LEHR) parallel-beam collimator to image various phantoms based on Monte Carlo simulation method. To generate labeled datasets, three digital phantoms are simulated using SIMIND, which are the whole body of human (Mdp WB), the human trunk (Mdp ECT) and the Jaszczak phantom consisting of multiple geometric structures. The detector orbit is circular covering 360° , and the radius of rotation is set to 15 cm. The parallel-collimated SPECT projection data for this simulation consists of 120 projection views in a 128-dimensional detector array with detector element size 2.2 mm. We use an 18% main energy window centered at 141 keV. The gamma photons within this energy window are considered as primary or first-order scattered photons. Moreover, we simulate 1.8×10^8 photon histories per projection view for each phantom to suppress the photon-flux fluctuation. The total numbers of photon counts detected in 120 projection views for three phantoms are presented in Table 1. Projection data at two specific view angles for three phantoms are further shown in Figure 4.

A total of 790 2D sinograms are extracted from the projection data of above three phantoms. Among them 625 sinograms with clear data distribution are selected to ensure that the detected projection data has good quality. The performance of proposed neural network at multiple levels of projection view angles is explored. The sparse-view sinograms are of size 60×128 , 40×128 and 30×128 , respectively. The full-view sinogram (120×128) is regarded as the reference image and the mapping from sinogram with the size of 60×128 , 40×128 , 30×128 to full-view sinogram is learned by LU-Net. The selected 625 sinograms are divided into training set and test set. The training set

contains 528 sinograms, accounting for 84.5%, and the test set contains 97 sinograms, accounting for 15.5%. The details of datasets are also summarized in Table 1.

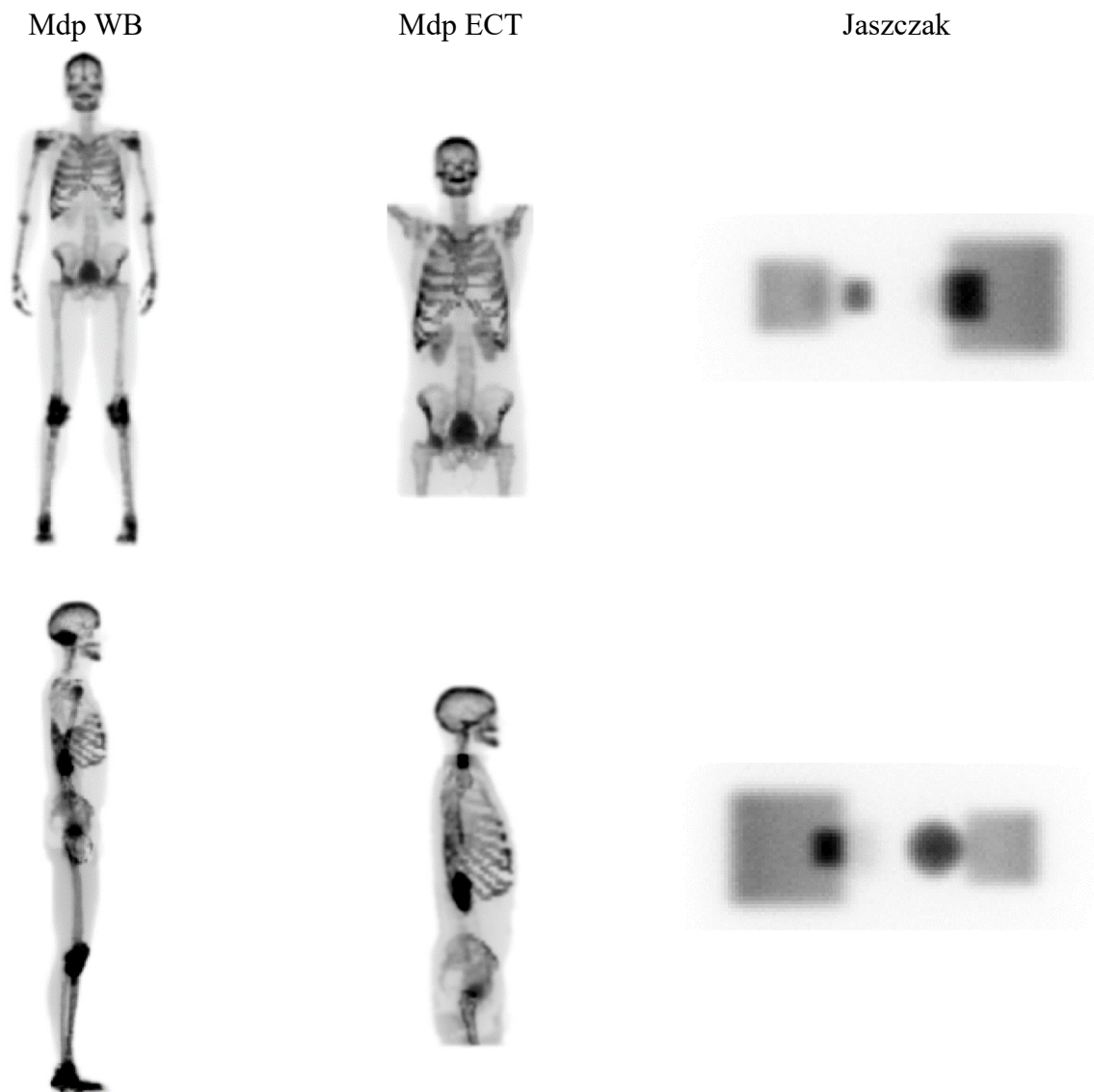


Figure 4. The projection data at coronal (top row) and sagittal (bottom row) view angles of Mdp WB, Mdp ECT and Jaszczak phantoms, respectively.

Table 1. Details of datasets.

Phantom	The size of 3D projection data	The total number of photon counts	The number of sinograms in training set	The number of sinograms in test set
Mdp WB	$128 \times 342 \times 120$	2.81×10^8	219	52
Mdp ECT	$128 \times 320 \times 120$	1.16×10^8	280	40
Jaszczak	$128 \times 128 \times 120$	8.30×10^8	29	5

3.3. Network training and implementation

The sparse-view sinograms are obtained by sampling the full-view sinograms in an interleaved fashion, during which one row of projection data corresponding to one view angle is collected in every few rows. Since multiple sparse-view levels (one-half, one-third and one-quarter) are analyzed in this paper, in order to maintain the unification and convenience, the size of sinogram for each sparse-view level is unified as 120×128 . The performance of sparse-view sinogram synthesis is mainly studied in this paper, therefore, interpolation algorithm is not employed to fill blank rows to avoid interference. Indeed, zeros are padded in the blank rows of sparse-view sinograms to maintain the size of full-view sinogram. With above procedure, sinograms with one-half, one-third and one-quarter of full view angles are generated. Example sinogram of each sparse-view level is shown in Figure 5.

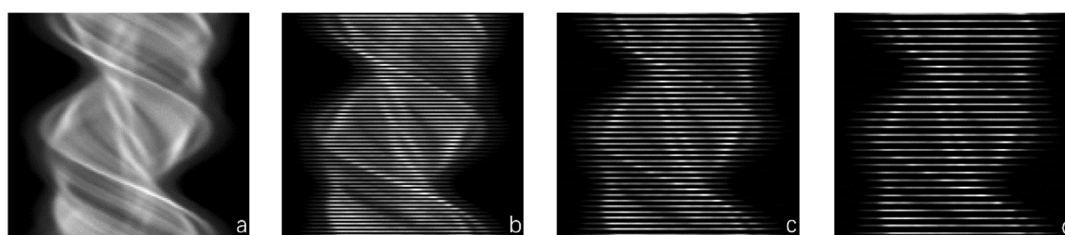


Figure 5. (a) Full-view (120 views) sinogram, (b) 60 views sinogram, (c) 40 views sinogram, (d) 30 views sinogram.

The proposed LU-Net is based on Pytorch neural network framework. Mean square error (MSE) function is used as loss function:

$$\mathcal{L}_{\text{MSE}} = \frac{1}{n} \sum_{i=1}^n (x_i - y_i)^2, \quad (11)$$

where x_i is the image processed by LU-Net, y_i is the label image and Adam optimizer is used to optimize loss function. Learning rate is set to 0.0001 and weight attenuation is 0.95. Network training is completed in NVIDIA Tesla P40 24GB graphical processing unit with training time of nearly 8 hours.

3.4. Quantification of reconstruction

To evaluate the performance of proposed neural network, we use three global image quality metrics such as normalized mean square error (NMSE), peak signal-to-noise ratio (PSNR), structural similarity (SSIM), and two local image quality metrics such as contrast recovery coefficient (CRC) and coefficient of variation (COV) to measure the quality of reconstructed images. Details of these performance metrics are described as follows.

3.4.1. Global image quality metrics

NMSE is a global image quality metric to evaluate the quality of reconstructed images. Its definition is as follows:

$$NMSE := \frac{\sum_{i=1}^N (f_i - g_i)^2}{\sum_{i=1}^N g_i^2}, \quad (12)$$

where f_i and g_i are the pixel activity of reconstructed image and the ground truth respectively, N is the number of all pixels in the image.

PSNR is an objective metric to measure the degree of image distortion or noise level. Its definition is as follows:

$$PSNR := 10 \cdot \log_{10} \left(\frac{MAX_I^2}{MSE} \right), \quad (13)$$

where MSE is the mean square error between reconstructed image and ground truth, I is the ground truth, MAX_I is possible maximum pixel value in the image.

SSIM is a commonly used metric to measure the similarity of two images in the field of medical images. Its definition is as follows:

$$SSIM(x, y) = \frac{(2\mu_x\mu_y + c_1)(2\sigma_{xy} + c_2)}{(\mu_x^2 + \mu_y^2 + c_1)(\sigma_x^2 + \sigma_y^2 + c_2)}, \quad (14)$$

where x and y are reconstructed image and ground-truth, respectively, μ_x is the average of x , μ_y is the average of y , σ_x^2 and σ_y^2 represent the variance of x and y , σ_{xy} is the covariance of x and y .

3.4.2. Local image quality metrics

CRC can be used to reflect the recovery of surrounding media, so it is often used in image processing tasks as a measure of local recovery, which is defined as follows [31]:

$$CRCmean_k = \frac{m_k}{f_k}, \quad (15)$$

where $k = 1, 2, \dots, n$ denotes the index of region of interest (ROI), m_k denotes the average of the k th ROI, and f_k is the true distribution of the k th ROI.

In image processing task, the coefficient of variation is usually used to represent the pixel-to-pixel variability in the image, which is defined as follows:

$$Cov_k = \frac{Std_k}{m_k}, \quad (16)$$

where Std_k and m_k represent standard deviation and average value of the k th ROI, respectively.

4. Results

4.1. Reconstructed image

The sinogram of each sparse-view level in test set is input into the trained LU-Net network, and the output is obtained as the synthesized sinogram. In Figure 6, we show the sinograms of chest slice from Mdp ECT phantom and those of Jaszczak phantom after LU-Net synthesis, respectively.

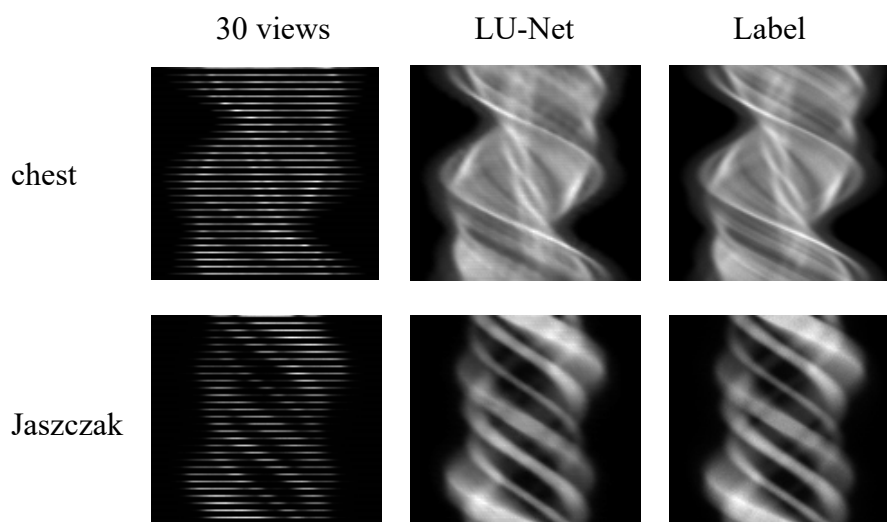


Figure 6. From left to right: 30 views sinogram, sinogram based on LU-Net, full-view sinogram.

The proposed method is compared with TV-PAPA and U-Net. Reconstructed images of the chest slice from Mdp ECT phantom are shown in Figure 7, where the label is the reconstructed image of 120 views(full-view) sinogram, the second column is the iterative reconstruction of 30 views sinogram using TV-PAPA, the latter two columns are reconstructed image results of 30 views sinogram synthesized by U-Net and LU-Net, respectively. In order to observe more intuitively, the contrast of yellow box is enhanced in Figure 7. The LU-Net shows superior performance in terms of preserving smooth regions, as compared to other methods. Both U-Net and LU-Net exhibit better staircase artifact reduction performance than traditional TV-PAPA. Furthermore, some global and local quantitative metrics of LU-Net are improved inordinately compared to U-Net structure. More details will be discussed later.

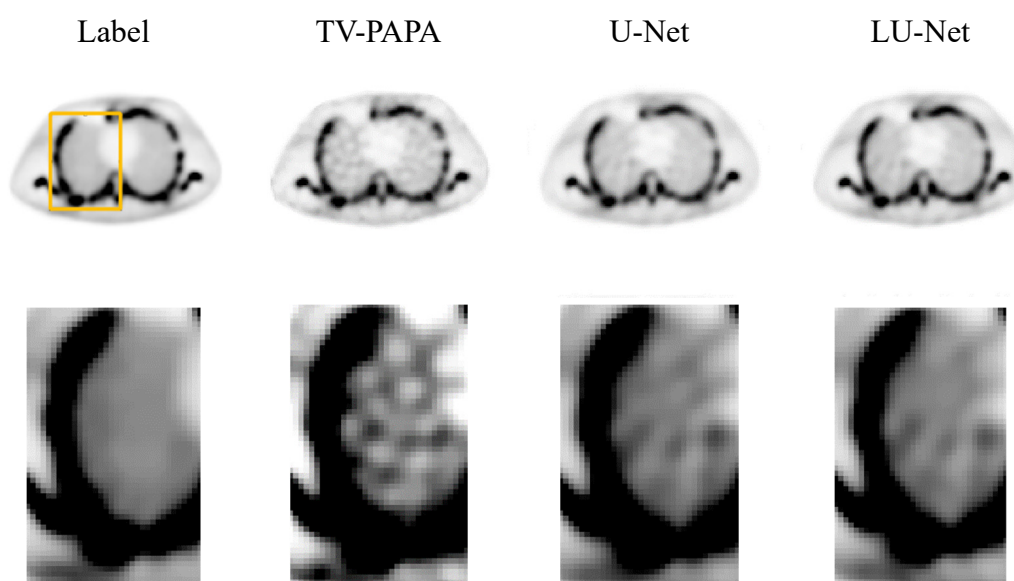


Figure 7. Reconstructed chest images. left to right: Image reconstruction of full-view sinogram, TV-PAPA with 30 views, U-Net with 30 views, LU-Net with 30 views.

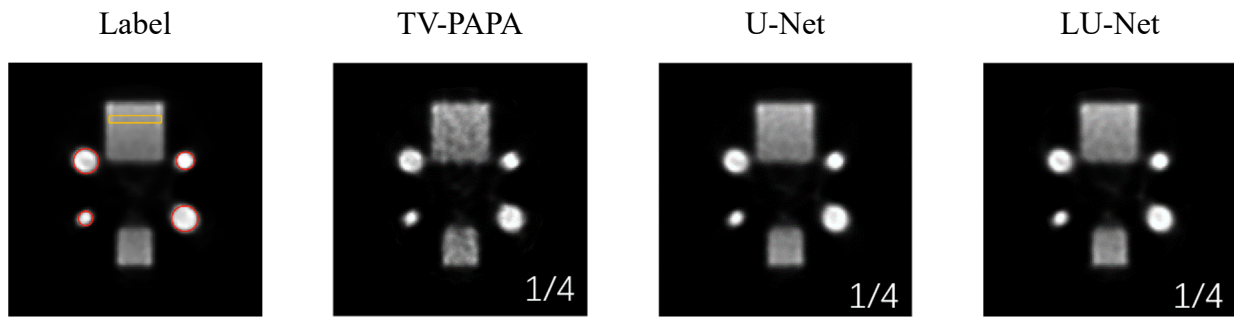


Figure 8. Reconstructed Jaszczak images. left to right: Image reconstruction of full-view sinogram, TV-PAPA with 30 views, U-Net with 30 views, LU-Net with 30 views.

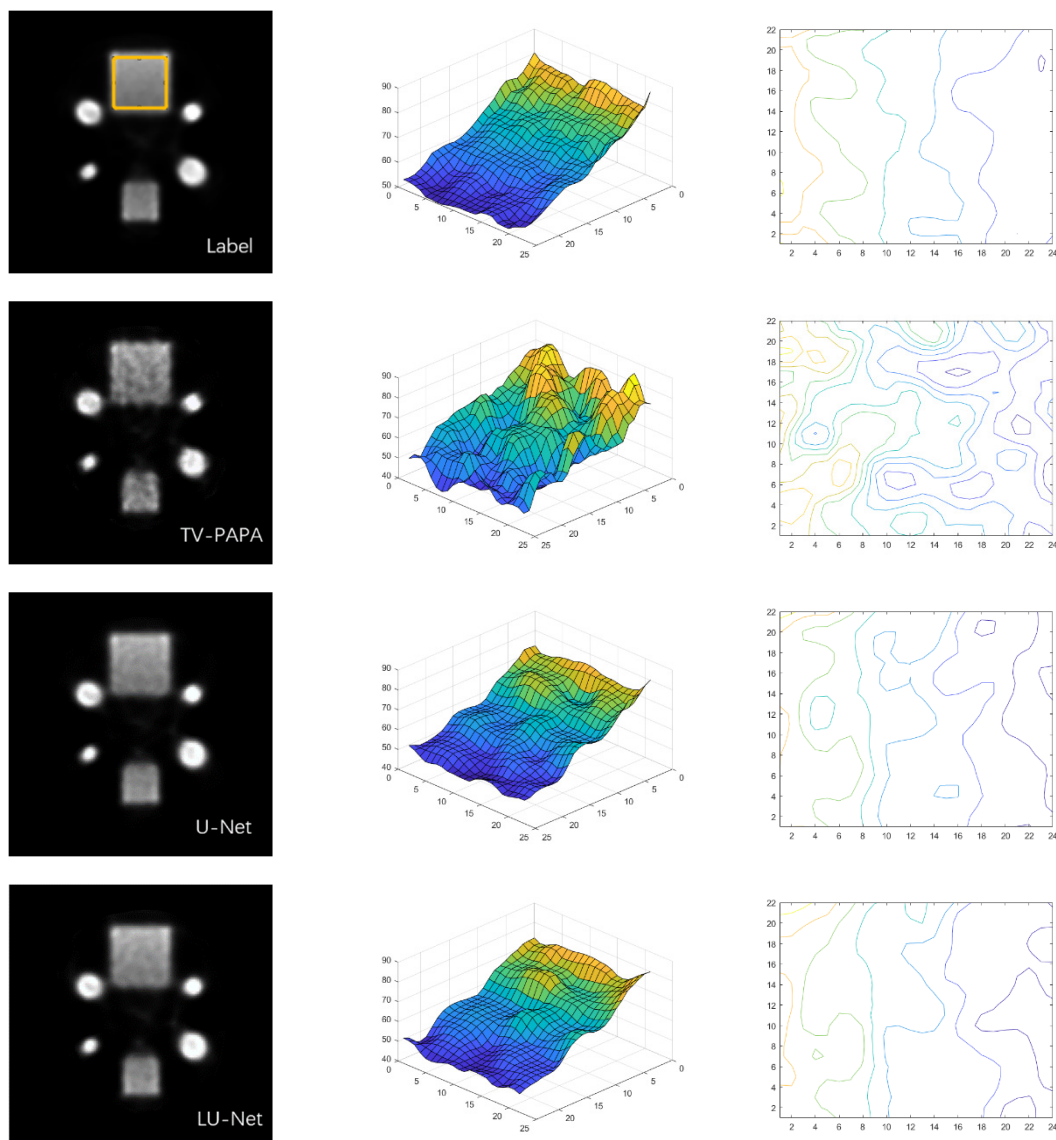


Figure 9. Surface and contour plots of reconstructed Jaszczak image.

Figure 8 shows the reconstructed image of a Jaszczak phantom. Yellow rectangular ROI is used for COV calculation and four red spherical ROIs are used for CRCmean calculation in section 4.2.2. To better show reconstruction performance, surface plot and corresponding contour plot of yellow rectangular ROI in each reconstructed Jaszczak image are shown in Figure 9. The less the artifacts and noise of reconstructed image, the smoother the corresponding surface plot and the less the closed-loops of contour plot. In this study, since 120 views sinograms are regarded as label data, the surface plot of reconstructed image from label data is still slightly steep, and the corresponding contour plot may also have a closed loop. In Figure 9, it can be seen that there are a large number of closed-loops in contour plot of reconstructed image with TV-PAPA. Closed-loops of reconstructed image with U-Net have significantly been reduced. It is obvious that the proposed network further reduces closed-loops, indicating that LU-Net achieves better performance in suppressing image artifacts and noise.

4.2. Quality analysis of reconstructed image

Three metrics including NMSE, PSNR and SSIM are calculated respectively for reconstructed chest images to evaluate the global quality of reconstructed images. Meanwhile, specific ROIs are selected on the reconstructed Jaszczak image, as well as contrast recovery coefficient and variation coefficient of selected ROIs are calculated, respectively.

4.2.1. Global quality analysis of image reconstruction

In order to make the quantitative comparison more intuitive, NMSE, PSNR and SSIM of reconstructed images at different projection view levels are calculated. The results are summarized in Tables 2 to 4. As shown in Table 2, we calculate SSIM, PSNR, and NMSE of reconstructed images at 30 projection view with three methods: TV-PAPA, sinogram synthesis with U-Net, and sinogram synthesis with LU-Net. All global metrics are averaged over the test set for Jaszczak phantom as well as for the chest region of Mdp WB and Mdp ECT phantoms, respectively. In Table 2, it can be observed that the results with TV-PAPA are the lowest in terms of SSIM and PSNR, while U-Net performs better than TV-PAPA and LU-Net performs best among these three methods. For NMSE, the results with LU-Net are the lowest, while those with U-Net and TV-PAPA are significantly higher, furthermore, TV-PAPA performs worst. These results indicate that the proposed network can maintain high quality of reconstructed images in low dose SPECT reconstruction task. There are similar conclusions in Tables 3 and 4.

Table 2. Global quality metrics of images reconstructed by three competing methods from 30-view projection data.

	Methods	SSIM	NMSE	PSNR
Chest of WB and ECT	TV-PAPA	0.5643	0.01268	37.52
	U-Net	0.8928	0.01141	37.91
	LU-Net	0.9293	0.00823	39.33
Jaszczak	TV-PAPA	0.5359	0.00889	36.88
	U-Net	0.8442	0.00612	38.50
	LU-Net	0.9117	0.00593	38.70

Table 3. Global quality metrics of images reconstructed by three competing methods from 40-view projection data.

	Methods	SSIM	NMSE	PSNR
Chest of WB and ECT	TV-PAPA	0.6888	0.00666	40.27
	U-Net	0.9032	0.00539	41.14
	LU-Net	0.9573	0.00476	41.73
Jaszczak	TV-PAPA	0.6537	0.00859	37.03
	U-Net	0.8574	0.00550	39.02
	LU-Net	0.9375	0.00472	39.63

Table 4. Global quality metrics of images reconstructed by three competing methods from 60-view projection data.

	Methods	SSIM	NMSE	PSNR
Chest of WB and ECT	TV-PAPA	0.8443	0.00300	43.90
	U-Net	0.9115	0.00219	45.04
	LU-Net	0.9691	0.00146	46.81
Jaszczak	TV-PAPA	0.8100	0.00340	41.16
	U-Net	0.8685	0.00193	43.51
	LU-Net	0.9470	0.00175	43.93

4.2.2. Local quality analysis of image reconstruction

To further evaluate local performance of proposed network in image reconstruction, five ROIs are selected in each image reconstructed by TV-PAPA, U-Net, and LU-Net, respectively. Five different reconstructed cross-sectional images of the Jaszczak phantom are selected in each of the three competing methods.

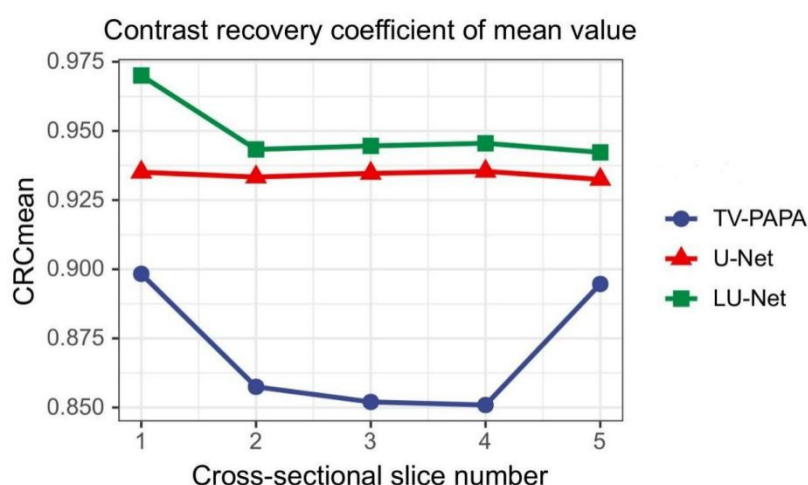


Figure 10. Average CRCmean over four circular ROIs in each reconstructed cross-sectional image of the Jaszczak phantom. The x-axis represents image number, and y-axis represents average CRCmean.

Four spherical ROIs are selected in Figure 8, which are marked in red for CRCmean calculation. CRCmean of each spherical ROI is calculated and average value over four CRCmeans for each reconstructed cross-sectional slice is shown in Figure 10. It can be observed that CRCmean of each reconstructed slice with TV-PAPA is much lower than that with U-Net and LU-Net. Furthermore, LU-Net consistently outperforms U-Net on CRCmeans of all five reconstructed slices.

Meanwhile, specific ROI is selected for reconstructed Jaszczak image at three projection view angle levels respectively, and coefficient of variation of ROI is calculated. Selected ROI is shown in Figure 8, which is marked in yellow. The results are summarized in Table 5, which indicate that the COV of reconstructed image with TV-PAPA is worse than those of other two methods in each projection view angle level, and the COV of reconstructed image with LU-Net is the best in each projection view angle level.

Table 5. Coefficients of variation for three competing reconstruction methods.

Method	30-view	40-view	60-view
TV-PAPA	9.34%	7.24%	6.03%
U-Net	7.47%	4.78%	4.69%
LU-Net	4.56%	4.33%	3.89%

4.3. Test results of different phantoms

In order to verify the generalization performance of proposed LU-Net, we further introduce three 2D phantoms with different statistics from the training data, which are shown in Figure 11. The parallel-collimated projection data of these three test phantoms are generated using SIMIND, all of which consist of 120 projection views in the 128-dimensional detector array. The total numbers of photon counts detected in 120 views for Brain-Lesion, Lumpy and Brain-HighResolution phantoms are 3.99×10^5 , 2.78×10^5 and 4.41×10^4 , respectively. The simulated full-view (120 views) data is then evenly sampled to generate sparse-view (30 views) sinograms for data synthesis. Three competing reconstruction methods are applied to above sparse-view test data for comparison. We calculate global metrics SSIM and PSNR of reconstructed images to evaluate reconstruction accuracy of each method on the test data with different statistics. The results are summarized in Table 6. This simulation shows that for test phantoms with different statistics from the training dataset, learning-based sinogram synthesis methods still outperform the traditional iterative reconstruction method TV-PAPA in terms of reconstruction accuracy. The LU-Net architecture proposed in this work performs the best in this category.

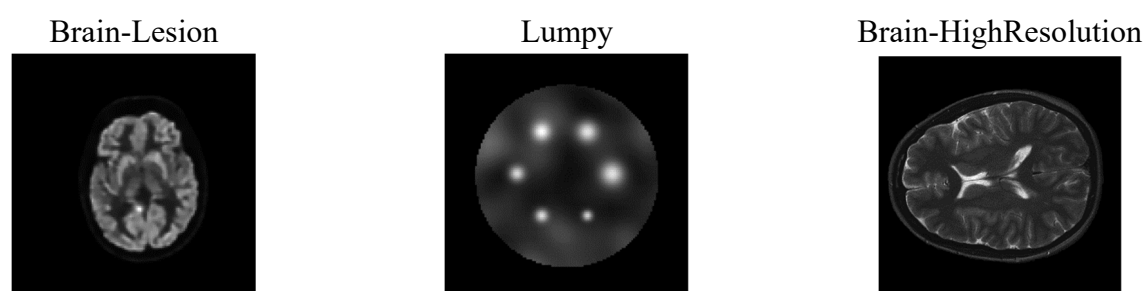


Figure 11. Three 2D phantoms with different statistics from training data.

Table 6. PSNR and SSIM of images reconstructed by three competing methods from 30-view projection data.

Phantom	Method	PSNR	SSIM
Brain-Lesion	TV-PAPA	16.12	0.7390
	U-Net	26.70	0.8246
	LU-Net	30.74	0.8345
Lumpy	TV-PAPA	18.21	0.7989
	U-Net	29.07	0.8320
	LU-Net	35.66	0.8539
Brain-HighResolution	TV-PAPA	16.88	0.7971
	U-Net	25.88	0.8213
	LU-Net	29.93	0.8419

4.4. Comparison between sparse-view and reduced-photon count

Table 7. PSNR and SSIM of images reconstructed with sparse-view and low-photon count sinograms.

Phantom	Method	PSNR	SSIM
Chest of WB and ECT	U-Net _{sparse-view}	41.94	0.8926
	U-Net _{low-count}	42.02	0.9083
	LU-Net _{sparse-view}	44.17	0.9492
	LU-Net _{low-count}	44.23	0.9514
Jaszczak	U-Net _{sparse-view}	40.01	0.8510
	U-Net _{low-count}	40.05	0.8573
	LU-Net _{sparse-view}	41.75	0.8820
	LU-Net _{low-count}	41.83	0.8823

In SPECT imaging, lower radiation dose can be achieved by reducing the radiotracer activity, which inevitably leads to projection data with either sparse projection views or reduced photon counts per view. In order to evaluate the feasibility of both low-dose cases with learning-based methodology, we generate two sets of projection data and perform network training in respective case. Indeed, a SIEMENS E.CAM gamma camera with LEHR parallel-beam collimator is simulated. The parallel-collimated SPECT projection data in this simulation consists of 120 projection views in a 128-dimensional detector array with detector element size 2.2 mm and is generated using SIMIND. The first projection dataset corresponds to sparse-view (60 views) case with full-count per view, which is generated via evenly sampling the full-view data. The second projection dataset corresponds to full-view case with reduced-count per view, which is obtained via dividing the number of photon counts in original simulated full-view data by 2. The comparable low-count levels of sinogram are realized by means of both ways. Indeed, the total numbers of photon counts detected in 60 projection views (or in 120 views with one-half counts per view) for Mdp WB, Mdp ECT and Jaszczak phantoms are 1.41×10^8 , 5.81×10^7 and 4.15×10^7 , respectively. Based on above noise-free projection data, we implement Poisson noise for each projection dataset (i.e., sparse-view projection data with full-count per view and full-view projection data with reduced-count per view). We follow the same procedure

in Section 3.2 to prepare training and test datasets, and perform LU-Net and U-Net training in respective case. SSIM and PSNR of reconstructed images in test dataset are calculated and summarized in Table 7. This evaluation shows that sparse-view sampling and reduced-photon count collection are comparable approaches to achieve low-dose imaging, in some sense, both approaches produce reconstructed images of comparable quality at low-count level. Moreover, the proposed LU-Net performs better than U-Net in every aspect.

5. Conclusions

In this study, the synthesis performance of LU-Net for sparse-view projection data at various projection view levels (1/2, 1/3, 1/4 of full-view) is explored, as well as compared with traditional TV-PAPA and U-Net. The experimental results show that learning-based method is superior to traditional iterative reconstruction method. Compared with conventional U-Net, LU-Net performs better in global and local metrics, indicating that the missing views of sinogram may be better synthesized after considering the sequence feature of projection data at adjacent projection view angles. At the same time, the results also show that the proposed network architecture performs better in suppressing noise and image artifacts. Overall, this study shows that the proposed network architecture has the potential to reduce the dose of radiotracers required by SPECT imaging without compromising the reconstructed image quality.

Acknowledgments

This work is supported in part by the Natural Science Foundation of China under grants 11771464 and 11601537, and by the Natural Science Foundation of Guangdong Province under grant 2021A1515012290, and by the Opening Project of Guangdong Province Key Laboratory of Computational Science at Sun Yat-sen University under grant 2021007, and by the Science and Technology Program of Guangzhou under grant 201804020053, and by Guangdong Provincial Key Laboratory of Cyber-Physical Systems under grant 2020B1212060069, and by National & Local Joint Engineering Research Center of Intelligent Manufacturing Cyber-Physical Systems.

Conflict of interest

The authors have no relevant financial or non-financial interests to disclose.

References

1. H. A. Ziessman, J. P. O'Malley, J. H. Thrall, *Nuclear Medicine: The Requisites*, Mosby, 2006.
2. R. G. Wells, Dose reduction is good but it is image quality that matters, *J. Nucl. Cardiol.*, **27** (2020), 238–240. <https://doi.org/10.1007/s12350-018-1378-5>
3. J. Zhang, S. Li, A. Krol, C. R. Schmidtlein, E. Lipson, D. Feiglin, et al., Infimal convolution-based regularization for SPECT reconstruction, *Med. Phys.*, **45** (2018), 5397–5410. <https://doi.org/10.1002/mp.13226>

4. A. Krol, S. Li, L. Shen, Y. Xu, Preconditioned alternating projection algorithms for maximum a posteriori ECT reconstruction, *Inverse Probl.*, **28** (2012). <https://doi.org/10.1088/0266-5611/28/11/115005>
5. Y. Jiang, S. Li, S. X. Xu, A higher-order polynomial method for SPECT reconstruction, *IEEE Trans. Med. Imaging*, **38** (2019), 1271–1283. <https://doi.org/10.1109/TMI.2018.2881919>
6. M. H. Zhang, B. Dong, A review on deep learning in medical image reconstruction, *J. Oper. Res. Soc. China*, **8** (2020), 311–340. <https://doi.org/10.1007/s40305-019-00287-4>
7. A. J. Ramon, Y. Yang, P. H. Pretorius, K. L. Johnson, M. A. King, M. N. Wernick, Improving diagnostic accuracy in low-dose SPECT myocardial perfusion imaging with convolutional denoising networks, *IEEE Trans. Med. Imaging*, **39** (2020), 2893–2903. <https://doi.org/10.1109/TMI.2020.2979940>
8. H. Chen, Y. Zhang, M. K. Kalra, F. Lin, Y. Chen, P. Liao, et al., Low-dose CT with a residual encoder-decoder convolutional neural network, *IEEE Trans. Med. Imaging*, **36** (2017), 2524–2535. <https://doi.org/10.1109/TMI.2017.2715284>
9. C. Z. Zhang, K. X. Liang, X. Dong, Y. Q. Xie, G. H. Cao, A sparse-view CT reconstruction method based on combination of DenseNet and deconvolution, *IEEE Trans. Med. Imaging*, **37** (2018), 1407–1477. <https://doi.org/10.1109/TMI.2018.2823338>
10. Q. Yang, P. Yan, Y. Zhang, H. Yu, Y. Shi, X. Mou, et al., Low-dose CT image denoising using a generative adversarial network with Wasserstein distance and perceptual loss, *IEEE Trans. Med. Imaging*, **37** (2018), 1348–1357. <https://doi.org/10.1109/TMI.2018.2827462>
11. N. Bastys, V. Grau, Super resolution of cardiac cine MRI sequences using deep learning, *Image Analys. Moving Org. Breast Thorac. Images*, **2018** (2018), 23–31. https://doi.org/10.1007/978-3-030-00946-5_3
12. J. Sun, H. Li, Z. Xu, Deep ADMM-Net for compressive sensing MRI, *Adv. Neural Inf. Process. Syst.*, **2016** (2016), 29.
13. Y. Yang, J. Sun, H. Li, Z. Xu, ADMM-CSNet: a deep learning approach for image compressive sensing, *IEEE Trans. Pattern Anal. Mach. Intell.*, **42** (2018), 521–538. <https://doi.org/10.1109/TPAMI.2018.2883941>
14. J. Adler, O. Öktem, Learned primal-dual reconstruction, *IEEE Trans. Med. Imaging*, **37** (2018), 1322–1332. <https://doi.org/10.1109/TMI.2018.2799231>
15. M. H. Zhang, B. Dong, D. B. Liu, JSR-Net: a deep network for Joint Spatial-Radon domain CT reconstruction from incomplete data, in *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, (2019), 3657–3661. <https://doi.org/10.1109/ICASSP.2019.8682178>
16. I. Haeggstroem, C. R. Schmidlein, G. Campanella, T. J. Fuchs, DeepPET: a deep encoder-decoder network for directly solving the PET reconstruction inverse problem, *Med. Image Anal.*, **54** (2018), 253–262. <https://doi.org/10.1016/j.media.2019.03.013>
17. X. Dong, S. Vekhande, G. Cao, Sinogram interpolation for sparse-view micro-CT with deep learning neural network, in *Medical Imaging 2019: Physics of Medical Imaging*, SPIE, (2019), 692–698. <https://doi.org/10.1117/12.2512979>
18. H. Yuan, J. Jia, Z. Zhu, SIPID: a deep learning framework for sinogram interpolation and image denoising in low-dose CT reconstruction, in *2018 IEEE 15th International Symposium on Biomedical Imaging*, (2018), 1521–1524. <https://doi.org/10.1109/ISBI.2018.8363862>

19. H. Lee, J. Lee, H. Kin, B. Cho, S. Cho, Deep-neural-network based sinogram synthesis for sparse-view CT image reconstruction, *IEEE Trans. Radiat. Plasma Med. Sci.*, **3** (2019), 109–119. <https://doi.org/10.1109/TRPMS.2018.2867611>
20. I. Shiri, P. Sheikhzadeh, M. R. Ay, Deep-fill: deep learning based sinogram domain gap filling in positron emission tomography, preprint, arXiv: 1906.07168.
21. C. Tang, W. Zhang, L. Wang, A. Cai, N. Liang, B. Yang, Generative adversarial network-based sinogram super-resolution for computed tomography imaging, *Phys. Med. Biol.*, **65** (2020), 235006.
22. Z. Li, W. Zhang, L. Wang, A. Cai, N. Liang, B. Yang, et al., A sinogram inpainting method based on generative adversarial network for limited-angle computed tomography, in *15th International Meeting on Fully Three-Dimensional Image Reconstruction in Radiology and Nuclear Medicine. International Society for Optics and Photonics*, (2019), 11072. <https://doi.org/10.1117/12.2533757>
23. Y. Wang, W. Zhang, A. Cai, L. Wang, C. Tang, Z. Feng, et al., An effective sinogram inpainting for complementary limited-angle dual-energy computed tomography imaging using generative adversarial networks, *J. X-Ray Sci. Technol.*, **29** (2020), 1–25. <https://doi.org/10.3233/XST-200736>
24. C. Chrysostomou, L. Koutsantonis, C. Lemesios, C. N. Papanicolas, SPECT angle interpolation based on deep learning methodologies, in *2020 IEEE Nuclear Science Symposium and Medical Imaging Conference (NSS/MIC)*, (2021), 1–4. <https://doi.org/10.1109/NSS/MIC42677.2020.9507966>
25. O. Ronneberger, P. Fischer, T. Brox, U-net: convolutional networks for biomedical image segmentation, in *International Conference on Medical Image Computing and Computer-Assisted Intervention*, (2015), 234–241. https://doi.org/10.1007/978-3-319-24574-4_28
26. M. T. McCann, K. H. Jin, M. Unser, Deep convolutional neural network for inverse problems in imaging, *IEEE Trans. Image Process.*, **34** (2017), 85–95. <https://doi.org/10.1109/MSP.2017.2739299>
27. Y. S. Han, J. Yoo, J. C. Ye, Deep residual learning for compressed sensing CT reconstruction via persistent homology analysis, preprint, arXiv:1611.06391.
28. S. Hochreiter, J. Schmidhuber, Long short-term memory, *Neural Comput.*, **9** (1997), 1735–1780. <https://doi.org/10.1162/neco.1997.9.8.1735>
29. L. A. Shepp, Y. Vardi, Maximum likelihood reconstruction for emission tomography, *IEEE Trans. Med. Imaging*, **1** (1982), 113–122. <https://doi.org/10.1109/TMI.1982.4307558>
30. C. A. Micchelli, X. L. Shen, S. Y. Xu, Proximity algorithms for image models: denoising, *Inverse Probl.*, **27** (2011), 045009.
31. S. Tong, A. M. Alessio, P. E. Kinahan, Noise and signal properties in PSF-based fully 3D PET image reconstruction: an experimental evaluation, *Phys. Med. Biol.*, **55** (2016), 1453–1473.



AIMS Press

©2022 the Author(s), licensee AIMS Press. This is an open access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/4.0>)