

Research Article

Wenqi Ao, Wenbin Li* and Jianliang Qian

A data and knowledge driven approach for SPECT using convolutional neural networks and iterative algorithms

<https://doi.org/10.1515/jiip-2020-0056>

Received May 17, 2020; accepted January 3, 2021

Abstract: We propose a data and knowledge driven approach for SPECT by combining a classical iterative algorithm of SPECT with a convolutional neural network. The classical iterative algorithm, such as ART and ML-EM, is employed to provide the model knowledge of SPECT. A modified U-net is then connected to exploit further features of reconstructed images and data sinograms of SPECT. We provide mathematical formulations for the architecture of the proposed networks. The networks are trained by supervised learning using the technique of mini-batch optimization. We apply the trained networks to the problems of simulated lung perfusion imaging and simulated myocardial perfusion imaging, and numerical results demonstrate their effectiveness of reconstructing source images from noisy data measurements.

Keywords: SPECT, data and knowledge driven, convolutional neural networks, iterative algorithms

MSC 2010: 65R32, 92C55

1 Introduction

The single photon emission computerized tomography (SPECT) is a nuclear medicine imaging technique designed to visualize the functional level of a part of the body. SPECT works in the following way: a radioactive chemical is injected into the bloodstream, which accumulates in a place of interest in the body, and emitted gamma rays attenuated by the body are detected outside. The measurements of attenuated rays can be modeled by the attenuated Radon transform

$$\mathcal{R}_a f(x, \theta) = \int_{\mathbb{R}} f(x + t\theta) e^{-\mathcal{B}a(x+t\theta, \theta)} dt, \quad x \in \mathbb{R}^2, \theta \in S^1 \quad (1.1)$$

where $\mathcal{B}a$ denotes the beam transform of a , $\mathcal{B}a(x, \theta) = \int_0^\infty a(x + \tau\theta) d\tau$. In equation (1.1), $a(x)$ is the attenuation function due to the mass density of tissue, and $f(x)$ is the source function that reveals the concentration of the radioactive chemical within the body. The main task of SPECT is to reconstruct the source function $f(x)$ from data measurements $\mathcal{R}_a f(x, \theta)$.

In typical applications of SPECT, the attenuation function $a(x)$ is assumed to be known, e.g., from an additional CT scan. In this setup, $f(x)$ can be reconstructed directly by means of explicit formulas. The first analytic reconstruction method is developed in [3], and the first Radon-type explicit inversion formula is given in [18]. The filtered back-projection (FBP) algorithm is the representative direct inversion method that is widely used in medical imaging. The direct inversion approach is fast and has proven convergence [8],

*Corresponding author: Wenbin Li, School of Science, Harbin Institute of Technology, Shenzhen, Shenzhen 518055, China, e-mail: liwenbin@hit.edu.cn

Wenqi Ao, School of Science, Harbin Institute of Technology, Shenzhen, Shenzhen 518055, China, e-mail: liwenbin@hit.edu.cn

Jianliang Qian, Department of Mathematics, Michigan State University, East Lansing, MI 48824, USA, e-mail: qian@math.msu.edu

but it is sensitive to noise so that the inversion result relies heavily on the completeness and high quality of measurement data. Problems of this type are more amenable to solution by iterative algorithms.

The attenuated Radon transform is linear with respect to $f(x)$ if $a(x)$ is known, and an iterative reconstruction algorithm solves the system of linear equations after discretization. In computerized tomography, the algebraic reconstruction technique (ART) [9] is widely used, which reconstructs the solution from a series of projections and is known as Kaczmarz method in numerical linear algebra. In SPECT, statistical iterative reconstruction algorithms are favorable, and among them, the most successful ones are the maximum-likelihood expectation maximization (ML-EM) algorithm [21], the ordered subset expectation maximization (OS-EM) algorithm [11], and the maximum a posteriori (MAP) reconstruction method [16]. The iterative algorithms can include additional regularization terms, such as l^1 or total variation regularization [4, 22], to enhance sparsity or sharp edges of the reconstructed solution. The tomography problem is then solved by minimizing an objective function consisting of data discrepancies and regularization terms, and optimization algorithms such as the gradient descent method, the ADMM method [23], and the primal-dual method [24] can be used in this scenario.

In recent years, utilizing deep learning neural networks has arisen as a new trend in imaging science. The deep learning methods have provided overwhelming performance for image segmentation [6, 19], classification [15], and recognition [10, 20]. The convolutional neural network architecture plays a key role in state-of-the-art deep learning approaches. Comparing to traditional fully connected networks, the convolutional neural network has a lower scale of connectedness and complexity, which makes it less prone to over-fitting. Therefore, researchers are beginning to embrace the use of convolutional neural networks in medical image reconstructions. For example, [1] proposes a partially learned gradient descent approach for computerized tomography reconstruction, where the convolutional network is employed to learn the gradient component in each iteration; [2] further proposes a learned primal-dual algorithm for tomographic reconstruction, where the convolutional network is used to replace proximal operators in the primal-dual optimization method; [13] combines the filtered back-projection (FBP) with the convolutional neural network to solve normal-convolutional inverse problems, such as the computerized tomography reconstruction; [7] uses the same idea of [13] to accelerate image reconstructions of SPECT. In the above works, both the deep learning approaches and the classical iterative or direct inversion algorithms are employed when developing reconstruction methods. In [1, 2], the convolutional network is incorporated into the iterative optimization algorithms; in [7, 13], the direct-inversion FBP algorithm is applied before using the neural network learning approach. While classical iterative or direct inversion algorithms utilize the knowledge of tomographic models which captures the mathematical relation between reconstructed objects and measurement data, convolutional neural networks are able to exploit hidden features of the input data by providing a sophisticated architecture between its input and output. As a result, the combination of a learning approach and a classical inversion algorithm leads to a data and knowledge driven approach that is able to use both features of data and knowledge of models.

Inspired by the works of [1, 2, 7, 13], we propose a reconstruction method for SPECT using convolutional neural networks and iterative algorithms. The basic structure of the proposed method is to apply the classical iterative algorithms to the measurement data at first, and then use the result as the input of a convolutional neural network which is trained to refine the solution to the ground truth image, yielding a data and knowledge driven approach. In the knowledge driven part, only a small number of iterations are performed when using the classical iterative algorithms. A rough prediction of the ground truth is adequate to exploit the knowledge of tomographic models. In the deep learning part, we propose a modified U-net architecture to link the prediction of the iterative algorithm to the final reconstructed image. The U-net is a convolutional neural network that was originally designed for biomedical image segmentation [19], and it is able to exploit more precise features with fewer training samples. In this work, we only reconstruct the source function $f(x)$ by assuming that the attenuation coefficient $a(x)$ is known. The framework of the proposed method can be extended to solve the identification problem of SPECT [17], where both the source and attenuation functions are to be determined, and it is an ongoing work.

The rest of the paper is organized as follows. In Section 2, we provide the classical iterative algorithms which will be used in this work. In Section 3, we develop the data and knowledge driven approach for SPECT, where both an ART U-net and an ML-EM U-net are developed. In Section 4, we apply the proposed method to simulated lung perfusion imaging and myocardial perfusion imaging. In Section 5, we draw our conclusion and discuss future works.

2 Iterative algorithms

The discretized version of SPECT is a linear system when $a(x)$ is known,

$$\mathbf{p} = \mathbf{A}\mathbf{f}, \quad (2.1)$$

where $\mathbf{f} = (f_j)_{N \times 1}$ denotes the vector of a source image, $\mathbf{p} = (p_i)_{M \times 1}$ simulates the measurement data in the sinogram, and $\mathbf{A} = (\omega_{i,j})_{M \times N}$ is the kernel matrix of the ray integral. The component $\omega_{i,j}$ represents the contribution of the j -th pixel of the source image to the i -th datum in the sinogram. We provide two classical iterative algorithms for (2.1) that will be used in this paper. The motivation is that even if the applied iterative algorithm does not yield a perfect fit to the data of the tomographic problem, feeding the imperfect reconstruction into the convolutional neural network will be able to achieve a decent overall reconstruction. As a result, we choose two simple and widely used iterative algorithms in tomographic reconstructions.

2.1 ART algorithm

The algebraic reconstruction technique (ART) computes the reconstruction image from a series of projections. Although it was intended for tomographic problems using transmission rays, such as traditional computerized tomography (CT), it can be applied to SPECT as well. Starting from an initial guess of the source image, such as $\mathbf{f}^{(0)} = \mathbf{0}$, the method updates $\mathbf{f}$ according to the following formula:

$$\mathbf{f}^{(k)} = \mathbf{f}^{(k-1)} + r_k \frac{p_i - \mathbf{a}_i \mathbf{f}^{(k-1)}}{\|\mathbf{a}_i\|_2^2} \mathbf{a}_i^T, \quad k \in \mathbb{N}^+, i = 1, 2, \dots, M, \quad (2.2)$$

where $\mathbf{a}_i$ denotes the i -th row of the matrix $\mathbf{A}$, $\mathbf{a}_i = (\omega_{i,1}, \dots, \omega_{i,N})$, and r_k is an optional relaxation parameter of the range $0 < r_k < 2$. We are not trying to adjust the value of r_k in this work, but simply set it to be 0.5 when applying the ART algorithm. One round of ART iterations implies one complete sweep of the measurement data p_i , where i goes from 1 to M .

2.2 ML-EM algorithm

The maximum-likelihood expectation maximization (ML-EM) algorithm calculates the most likely distribution of the solution from data measurements. The method is widely used for image reconstruction in the emission tomography such as PET and SPECT. We adopt a standard version of the ML-EM algorithm which uses the following formula to update $\mathbf{f}$:

$$f_j^{(k)} = \frac{f_j^{(k-1)}}{\sum_{i=1}^M \omega_{i,j}} \sum_{i=1}^M \omega_{i,j} \frac{p_i}{\sum_{j'=1}^N \omega_{i,j'} f_{j'}^{(k-1)}}, \quad 1 \leq j \leq N, k \in \mathbb{N}^+.$$

The initial guess of the source function can be taken as $\mathbf{f}^{(0)} = \mathbf{1}$, and the ML-EM algorithm ensures the positivity of $\mathbf{f}$ in iterations.

3 Proposed method: ART U-net and ML-EM U-net

3.1 Framework of the algorithm

We propose a data and knowledge driven approach for SPECT using both iterative algorithms and convolutional neural networks. Denote the measurement data by $\mathbf{p}^* \in \mathbb{R}^M$ and the true image by $\mathbf{f}^* \in \mathbb{R}^N$. We have the following model equation:

$$\mathbf{p}^* = \mathbf{A}\mathbf{f}^* + \delta\mathbf{p}, \quad (3.1)$$

where $\delta\mathbf{p}$ represents the contamination of noises.

Firstly, we establish an architecture that combines an iterative algorithm with a learning approach. Let $\mathcal{F}$ represent the operation of one round of iterations of a classical iterative algorithm,

$$\mathbf{f}^{(k)} = \mathcal{F}(\mathbf{p}^*; \mathbf{f}^{(k-1)}),$$

and Λ_Θ denotes the operation of a convolutional neural network, $\Lambda_\Theta: \mathbb{R}^N \rightarrow \mathbb{R}^N$, with Θ being the vector of network parameters to be estimated. The combined operation is written as the following:

$$\mathcal{A}_\Theta^\dagger = \Lambda_\Theta \circ \mathcal{F}^n, \quad (3.2)$$

where $\mathcal{F}^n = \mathcal{F} \circ \mathcal{F}^{n-1}$ and $\mathcal{F}^1 = \mathcal{F}$, so that $\mathcal{A}_\Theta^\dagger$ will provide an inverse operator for the SPECT model as shown in equation (3.1), where the approximate inverse operator is composed of the model-driven part captured by $\mathcal{F}^n$ and the data-driven part captured by Λ_Θ . We propose to use a modified U-net for constructing Λ_Θ and use the ART and ML-EM, respectively, for constructing $\mathcal{F}$. It leads to two combined networks for the solution of SPECT, which are named “ART U-net” and “ML-EM U-net”, and denoted by $\mathcal{A}_\Theta^{\dagger(\text{art})}$ and $\mathcal{A}_\Theta^{\dagger(\text{mlem})}$, respectively. In formula (3.2), the parameter n controls the number of iterations when using one of the classical iterative algorithms. For ART, $n = 1$ means a full round of iterations which uses the complete set of data measurements as shown in equation (2.2). The value of n must be small enough for efficiency and large enough to let the iterations work. Ideally, n should be the smallest number that allows the iterations to explore the tomographic model adequately.

We use a supervised learning approach to train the network $\mathcal{A}_\Theta^\dagger$ and determine the vector of parameters Θ . Let $\{(\mathbf{f}_s^*, \mathbf{p}_s^*) \mid s = 1, \dots, S\}$ be the set of training samples which includes simulated source images and data measurements. The loss function of one training sample is defined as the following:

$$L_s(\Theta) = \|\mathcal{A}_\Theta^\dagger(\mathbf{p}_s^*) - \mathbf{f}_s^*\|_2^2 + \lambda \|\Theta\|_2^2,$$

where the first term measures the misfit between the output solution and the sample image, and the second one provides an l^2 regularization for Θ aiming to suppress overfitting. The parameter λ controls the amount of regularization applied. The total cost function is then the average of the loss functions,

$$J(\Theta) = \frac{1}{S} \sum_{s=1}^S L_s(\Theta) = \frac{1}{S} \sum_{s=1}^S \|\mathcal{A}_\Theta^\dagger(\mathbf{p}_s^*) - \mathbf{f}_s^*\|_2^2 + \lambda \|\Theta\|_2^2,$$

and the vector Θ is determined by solving the optimization problem $\arg \min_\Theta J(\Theta)$.

We propose to use a mini-batch optimization technique which employs a small batch of the set of training samples at each iteration to update the parameter Θ . The mini-batch approach can achieve a balance between the efficiency and the robustness of training algorithms [5], and it is widely used for solving large-scale optimization problems in deep learning. Let n_0 denote the batch size so that the training samples are split into $N_0 = \frac{S}{n_0}$ batches,

$$\{(\mathbf{f}_s^*, \mathbf{p}_s^*) \mid s = 1, \dots, S\} = \bigcup_{t=1}^{N_0} \{(\mathbf{f}_s^*, \mathbf{p}_s^*) \mid s \in \Sigma_t\},$$

where $\Sigma_t = \{n_0(t-1) + 1, \dots, n_0 t\}$. Accordingly, the iteration update for Θ will be carried out based on the cost function for each batch,

$$J_t(\Theta) = \frac{1}{n_0} \sum_{s \in \Sigma_t} \|\mathcal{A}_\Theta^\dagger(\mathbf{p}_s^*) - \mathbf{f}_s^*\|_2^2 + \lambda \|\Theta\|_2^2. \quad (3.3)$$

The training vector Θ is updated according to the Adam algorithm [14], which is a state-of-the-art technique for the first-order gradient-based stochastic optimization. The formulas for updating Θ are given as follows:

$$\mathbf{g}^{(k)} = \nabla_{\Theta} J_t(\Theta^{(k-1)}), \quad (3.4)$$

$$\mathbf{m}^{(k)} = \beta_1 \mathbf{m}^{(k-1)} + (1 - \beta_1) \mathbf{g}^{(k)}, \quad (3.5)$$

$$\mathbf{v}^{(k)} = \beta_2 \mathbf{v}^{(k-1)} + (1 - \beta_2) (\mathbf{g}^{(k)})^2, \quad (3.6)$$

$$\hat{\mathbf{m}}^{(k)} = \frac{\mathbf{m}^{(k)}}{1 - \beta_1^k}, \quad (3.7)$$

$$\hat{\mathbf{v}}^{(k)} = \frac{\mathbf{v}^{(k)}}{1 - \beta_2^k}, \quad (3.8)$$

$$\Theta^{(k)} = \Theta^{(k-1)} - \alpha \frac{\hat{\mathbf{m}}^{(k)}}{\sqrt{\hat{\mathbf{v}}^{(k)} + \epsilon}}, \quad (3.9)$$

where k denotes the iteration number, $k \in \mathbb{N}^+$, and the batch number t is consecutively taken as $1, 2, \dots, N_0$, i.e. $t \equiv k \pmod{N_0}$. In formulas (3.4)–(3.9), the products and quotients between vectors indicate element-wise operations. The good default settings for the constant parameters are $\alpha = 0.001$, $\beta_1 = 0.9$, $\beta_2 = 0.999$, and $\epsilon = 10^{-8}$ (see [14]). In equation (3.4), the gradient of the cost function is evaluated in the following way:

$$\nabla_{\Theta} J_t(\Theta) = \frac{2}{m} \sum_{s \in \Sigma_t} (\nabla_{\Theta} \mathcal{A}_{\Theta}^{\dagger}(\mathbf{p}_s^*))^T (\mathcal{A}_{\Theta}^{\dagger}(\mathbf{p}_s^*) - \mathbf{f}_s^*) + 2\lambda \Theta,$$

where $\nabla_{\Theta} \mathcal{A}_{\Theta}^{\dagger}(\mathbf{p}_s^*) = \nabla_{\Theta} \Lambda_{\Theta}(\mathcal{F}^n(\mathbf{p}_s^*))$ and $\nabla_{\Theta} \Lambda_{\Theta}$ is evaluated by the back-propagation algorithm for supervised learning. Once the training of the parameter vector Θ is completed, the combined network $\mathcal{A}_{\Theta}^{\dagger} : \mathbb{R}^M \rightarrow \mathbb{R}^N$ is available, leading to

$$\mathbf{f}^{(\text{solution})} = \mathcal{A}_{\Theta}^{\dagger}(\mathbf{p}^*). \quad (3.10)$$

3.2 U-net for Λ_{Θ}

Recalling equations (3.2) and (3.10), the input of Λ_{Θ} is $\mathcal{F}^n(\mathbf{p}_s^*)$, which is an intermediate prediction of the source image, and the output of Λ_{Θ} is the refined reconstruction of the source image. We propose to use a modified U-net for constructing Λ_{Θ} .

The U-net is a convolutional neural network that was originally designed for biomedical image segmentation [19]. The network consists of a contracting path and a symmetric expanding path so that it has a U-shaped architecture. The contracting path is a typical convolutional network that consists of repeated applications of convolution, each followed by a rectified linear unit (ReLU) and a max pooling operation. The expanding path is designed to enable precise localization combined with contextual information from the contracting path. Every step in the expanding path consists of an upsampling followed by a convolution, a concatenation with the cropped feature map from the contracting path, and two convolutions, each followed by a ReLU. The U-net has the strength of exploiting more precise features with fewer training samples. In [13], some advantageous properties of the U-net for applications in computerized tomography are discussed.

Figure 1 shows the architecture of the modified U-net that is used in this work. Although the network has the same topology as the originally designed U-net in [19], there are some major differences. Firstly, the last layer leads to a single output image which expresses the source image, but the original U-net results in two channels: foreground and background. In addition, we add a sigmoid activation function so that the output value is between 0 and 1, which fits to features of the normalized source function taking values in $[0, 1]$. Secondly, since the input and output source images have the same size, we use zero-padding before each convolution so that the size of each feature map does not decrease. The concatenation operation consists of only copying instead of copying and cropping, as well. Thirdly, we add a batch normalization after each convolution, which is a state-of-the-art technique for accelerating training speed and improving the performance of optimization [12].

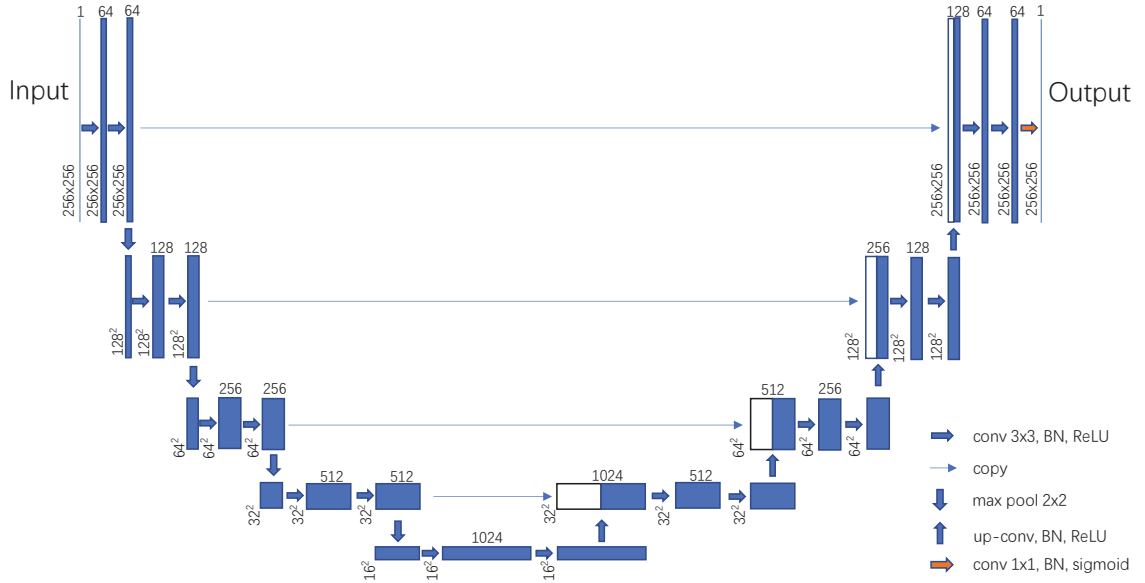


Figure 1: A modified U-net for the design of Λ_Θ .

4 Implementation and results

We apply the proposed method to the problems of simulated lung perfusion imaging and simulated myocardial perfusion imaging, respectively. In this section, we illustrate the reconstruction of source images with piecewise constant structures.

4.1 Data preparation

We will train the ART U-net and the ML-EM U-net in the context of lung perfusion imaging. The attenuation map is given in Figure 2, which is a CT image of lungs with 256×256 pixels. We prepare 1400 synthetic source images, denoted as $\{\mathbf{f}_s^* \mid s = 1, \dots, 1400\}$, and Figure 3 (a) shows six of them. Each source image consists of ellipses of random sizes and locations, and the resolution is uniformly of 256×256 pixels. The corresponding data sinograms, denoted as $\{\mathbf{p}_s^* \mid s = 1, \dots, 1400\}$, are of size 256 pixels by 60 views and are created by equation (3.1), where random noises are added to simulate the situation of practical measurements. To illustrate the capability of the combined network in fitting random noises, we use two types of noises and create two sets of data sinograms. Sinograms in the first set are contaminated by 5 % to 25 % Gaussian noises, which will be used for training, validation, and test of the ART U-net; Figure 3 (b) shows six of them. Sinograms in the second set are contaminated by 5 % to 25 % pepper and salt noises, which will be used for the ML-EM U-net, and Figure 3 (c) shows six of them.

4.2 Process of training

For each network ($\mathcal{A}_\Theta^{\dagger(\text{art})}$ or $\mathcal{A}_\Theta^{\dagger(\text{mlem})}$), we have 1400 pairs of samples including source images $\mathbf{f}_s^*$ and the corresponding sinograms $\mathbf{p}_s^*$. We separate the 1400 samples into three sets: the training set of 1000 samples, $\{(\mathbf{f}_s^*, \mathbf{p}_s^*) \mid s = 1, \dots, 1000\}$, the validation set of 200 samples, $\{(\mathbf{f}_s^*, \mathbf{p}_s^*) \mid s = 1001, \dots, 1200\}$, and the test set of the remaining 200 samples, $\{(\mathbf{f}_s^*, \mathbf{p}_s^*) \mid s = 1201, \dots, 1400\}$. To train the combined network as shown in equation (3.2), we must fix the hyperparameter n which indicates the number of iterations when employing the classical iterative algorithm. As discussed in Section 3.1, n should be small enough for efficiency and large enough to explore the model space adequately. In the context of supervised learning, such a hyperparameter



Figure 2: Attenuation map of lungs.

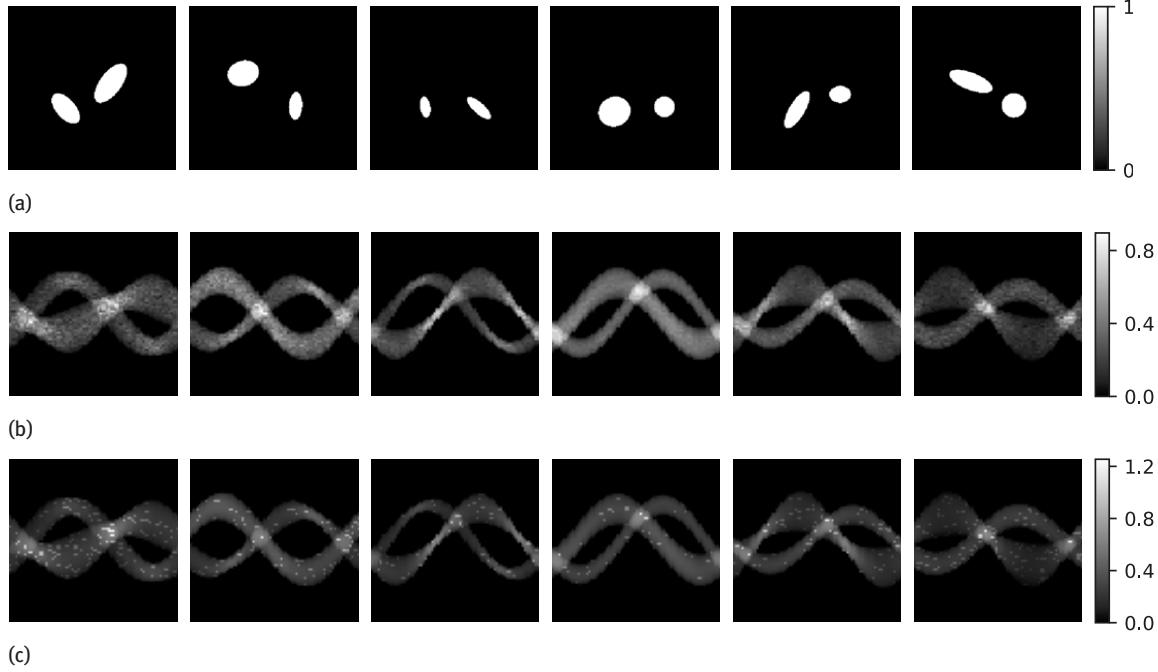


Figure 3: Six of the 1400 samples of source images and data sinograms. (a) Source images (256×256 pixels); (b) corresponding sinograms contaminated by 5 % to 25 % Gaussian noises; (c) corresponding sinograms contaminated by 5 % to 25 % pepper and salt noises. The sinograms are of size 256 pixels by 60 views.

can be tuned according to the performance of the trained network on the validation set. We take $n = 1$ for the ART U-net $\mathcal{A}_{\Theta}^{\dagger(\text{art})}$, which is the smallest number allowing ART to utilize all sinograms in the training set at least once. And for the ML-EM U-net $\mathcal{A}_{\Theta}^{\dagger(\text{mlem})}$, we take $n = 10$. We remark that just taking $n = 1$ for the ART will not guarantee that the iterative algorithm converges, and the same can be said when just taking $n = 10$ for the ML-EM. For example, Figure 4 shows the convergence history of the ART algorithm when applied to one of the prepared sinogram samples, where the discrepancy term used to evaluate the convergence of ART is defined as

$$D^{(k)} = \sqrt{\frac{1}{M} \sum_{i=1}^M \frac{|p_i - A_i \mathbf{f}^{(k)}|^2}{\|A_i\|_2^2}}$$

with A_i denoting the i -th row of the matrix $\mathbf{A}$ as shown in equation (2.1).

The combined network is trained on an Nvidia Tesla V100 GPU using the training set of 1000 samples. Meanwhile, the validation set of 200 samples is used to evaluate the training. The validation set does not contribute to the updating of Θ , but simply monitors the performance of the network in the training process. To implement the mini-batch optimization, we set the batch size n_0 to be $n_0 = 10$ so that 10 samples are employed in the network at each iteration. The regularization parameter in the mini-batch cost function $J_t(\Theta)$

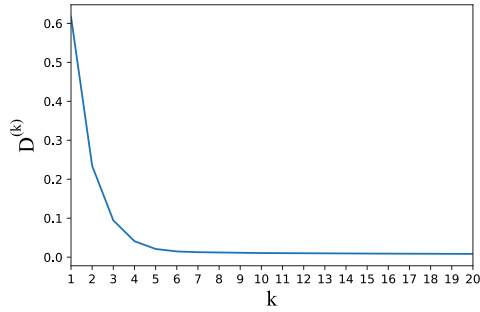


Figure 4: A convergence history of the ART algorithm. It is far from convergence at $k = 1$.

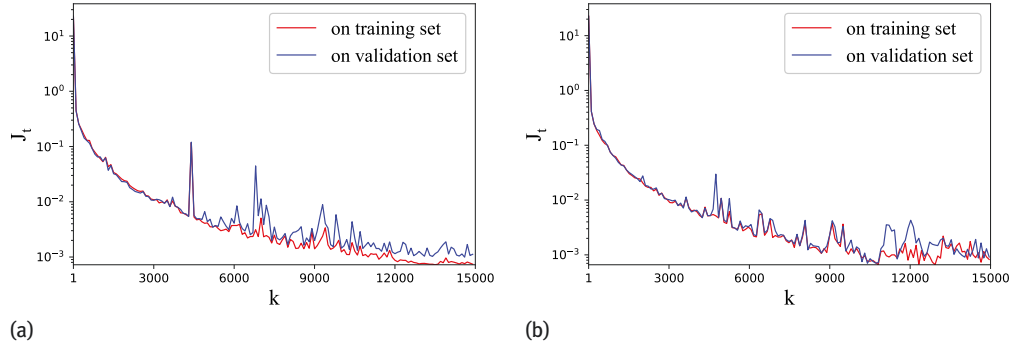


Figure 5: Convergence plots of the training process. (a) Convergence of $J_t(\Theta)$ for $\mathcal{A}_\Theta^{\dagger(\text{art})}$; (b) convergence of $J_t(\Theta)$ for $\mathcal{A}_\Theta^{\dagger(\text{mlem})}$.

(equation (3.3)) is taken to be $\lambda = 0.004$. Figure 5 (a) shows the convergence history of $J_t(\Theta)$ for the training of $\mathcal{A}_\Theta^{\dagger(\text{art})}$, and Figure 5 (b) shows the convergence history for the training of $\mathcal{A}_\Theta^{\dagger(\text{mlem})}$. In the convergence plots, the red line indicates the mini-batch cost function on the training set, and the blue line indicates the mini-batch cost function on the validation set.

Finally, we use the test set of 200 samples to evaluate the trained network. The mean squared error (MSE) on the source image is used to measure the performance of reconstructions, and it is defined in the following way:

$$\text{MSE} = \frac{1}{N} \|\mathbf{f}^* - \mathbf{f}^{(\text{solution})}\|_2^2, \quad \mathbf{f}^*, \mathbf{f}^{(\text{solution})} \in \mathbb{R}^N.$$

Figure 6 shows the values of MSE on the test samples, where Figure 6 (a) plots MSE for the trained ART U-net, and Figure 6 (b) plots MSE for the trained ML-EM U-net. We conclude that both networks are well trained and have decent performance on the samples from their corresponding test sets.

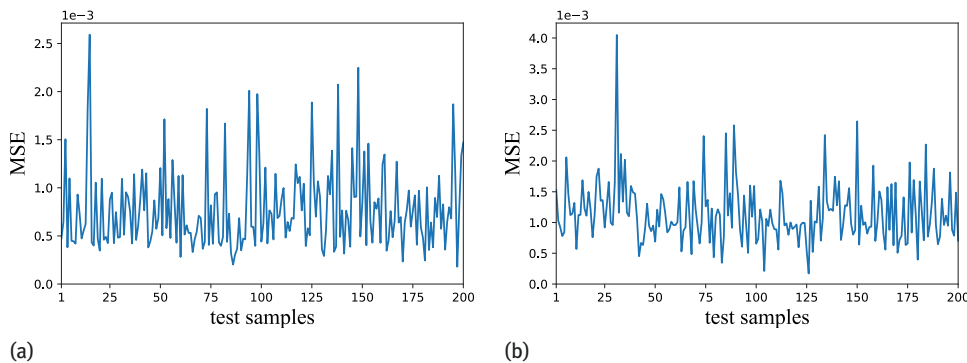


Figure 6: Mean squared error (MSE) of the trained network when applied to the test samples. (a) MSE for $\mathcal{A}_\Theta^{\dagger(\text{art})}$; (b) MSE for $\mathcal{A}_\Theta^{\dagger(\text{mlem})}$.

4.3 Reconstruction results

4.3.1 An example of lung perfusion imaging

The trained networks $\mathcal{A}_{\Theta}^{\dagger(\text{art})}$ and $\mathcal{A}_{\Theta}^{\dagger(\text{mlem})}$ are applied to the problem of simulated lung perfusion imaging. The attenuation map of lungs is given in Figure 2, which is assumed to be known in the reconstruction. Figure 7 (a) shows the exact source image, and Figure 7 (b) shows the clean sinogram data. The trained ART U-net $\mathcal{A}_{\Theta}^{\dagger(\text{art})}$ is applied to the clean sinogram and the sinograms contaminated by 10 % and 20 % Gaussian noises, respectively. Figure 8 shows the reconstruction results, where Figure 8 (a)–(c) plot the final solutions of the source image, and Figure 8 (d)–(f) plot the intermediate outputs of the ART iterations for illustration purpose. To quantitatively evaluate the performance of reconstructions, we compute the mean squared error (MSE) and the peak signal-to-noise ratio (PNSR), where a lower MSE and a higher PNSR correspond to a better reconstruction; Table 1 shows the results. The intermediate outputs of ART show fuzzy images with line artifacts, especially when Gaussian noises are added to the sinogram data. The combined ART U-net significantly refines the intermediate solutions and leads to correct reconstructions of the source image which has a piecewise constant structure.

Similarly, we apply the trained ML-EM U-net $\mathcal{A}_{\Theta}^{\dagger(\text{mlem})}$ to the clean sinogram data and the sinograms contaminated by 10 % and 20 % pepper and salt noises. Figure 9 shows the final solutions of the source image and the intermediate outputs of ML-EM. Table 2 provides the values of MSE and PNSR. Again, the trained network significantly improves the intermediate outputs of ML-EM and achieves much better reconstructions even when large amounts of pepper and salt noises are introduced.

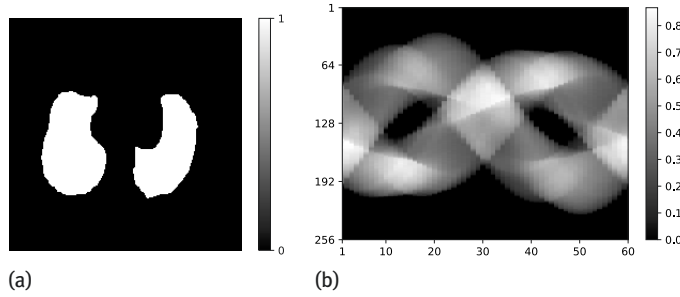


Figure 7: Lung perfusion imaging: source function and data sinogram.
(a) Exact source image; (b) clean data sinogram without noises.

Gaussian noises	MSE		PNSR	
	Final solution	Intermediate output	Final solution	Intermediate output
0	0.001781	0.009208	27.5	13.88
10 %	0.002417	0.018409	26.17	11.28
20 %	0.005447	0.046377	22.62	9.72

Table 1: Lung perfusion imaging: MSE and PNSR of the final solutions and the intermediate outputs of $\mathcal{A}_{\Theta}^{\dagger(\text{art})}$.

Pepper and salt noises	MSE		PNSR	
	Final solution	Intermediate output	Final solution	Intermediate output
0	0.005123	0.008485	22.90	19.76
10 %	0.004992	0.013466	23.01	12.11
20 %	0.004997	0.018745	23.01	11.08

Table 2: Lung perfusion imaging: MSE and PNSR of the final solutions and the intermediate outputs of $\mathcal{A}_{\Theta}^{\dagger(\text{mlem})}$.

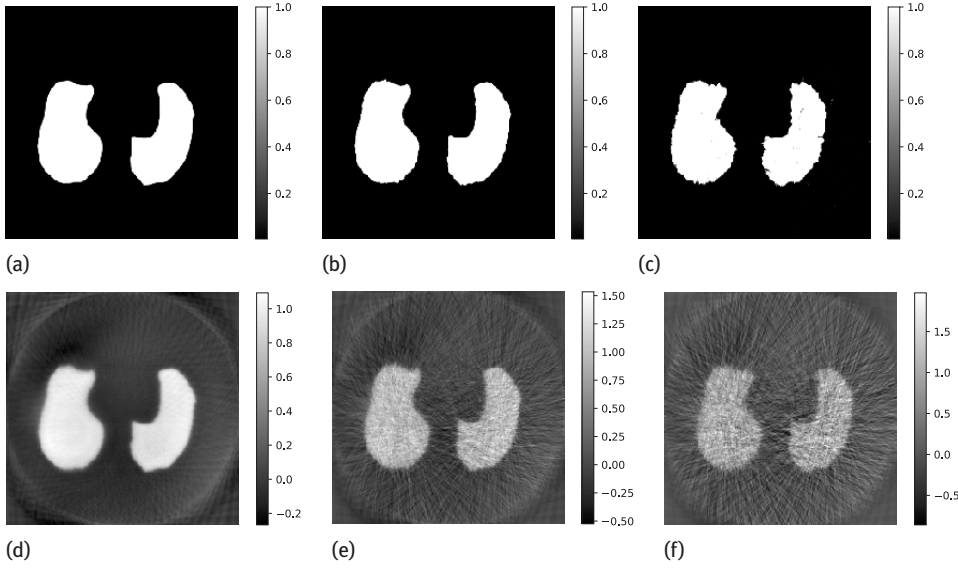


Figure 8: Lung perfusion imaging: solutions by $\mathcal{A}_\Theta^{\dagger(\text{art})}$. (a)–(c) Final solutions when applying $\mathcal{A}_\Theta^{\dagger(\text{art})}$ to the sinograms with 0, 10 %, and 20 % Gaussian noises; (d)–(f) intermediate outputs of ART in $\mathcal{A}_\Theta^{\dagger(\text{art})}$.

4.3.2 An example of myocardial perfusion imaging

We apply the trained networks $\mathcal{A}_\Theta^{\dagger(\text{art})}$ and $\mathcal{A}_\Theta^{\dagger(\text{mlem})}$ to the problem of simulated myocardial perfusion imaging. This example aims to illustrate the transferable feature of the proposed networks since they are only trained in the context of lung perfusion imaging.

Figure 10 (a) shows the attenuation map, which is a CT image of myocardium and is assumed to be known in the reconstruction. Figure 10 (b) shows the exact source image, and Figure 10 (c) shows the clean sinogram data.

The ART U-net $\mathcal{A}_\Theta^{\dagger(\text{art})}$ is applied to the clean sinogram and the sinograms contaminated by 10 % and 20 % Gaussian noises, respectively. Figure 11 shows the reconstruction results, where Figure 11 (a)–(c) plot the final solutions of the source image, and Figure 11 (d)–(f) plot the intermediate outputs of the ART iteration. Table 3 shows the values of MSE and PNSR for these solutions and intermediate outputs. The trained ART U-net successfully recovers the source image from noisy sinograms in the context of myocardial perfusion imaging. Although only one round of the ART iteration is employed and the intermediate outputs are contaminated by artifacts, the combined network is able to explore the feature of the source image, yielding good reconstructions even when 20 % Gaussian noises are introduced.

Similarly, the trained ML-EM U-net $\mathcal{A}_\Theta^{\dagger(\text{mlem})}$ is applied to the clean sinogram and the sinograms contaminated by 10 % and 20 % pepper and salt noises. The recovered solutions and intermediate outputs are shown in Figure 12; the values of MSE and PNSR for these results are provided in Table 4. We conclude that the reconstructions are successful, which recover very well the shape and concentration values of the source image. We conclude that the network trained in the lung perfusion imaging can be transferred to the problem of myocardial perfusion imaging.

Gaussian noises	MSE		PNSR	
	Final solution	Intermediate output	Final solution	Intermediate output
0	0.000402	0.002535	34.17	18.09
10 %	0.000528	0.003340	32.93	15.69
20 %	0.000854	0.005521	30.77	14.35

Table 3: Myocardial perfusion imaging: MSE and PNSR of the final solutions and the intermediate outputs of $\mathcal{A}_\Theta^{\dagger(\text{art})}$.

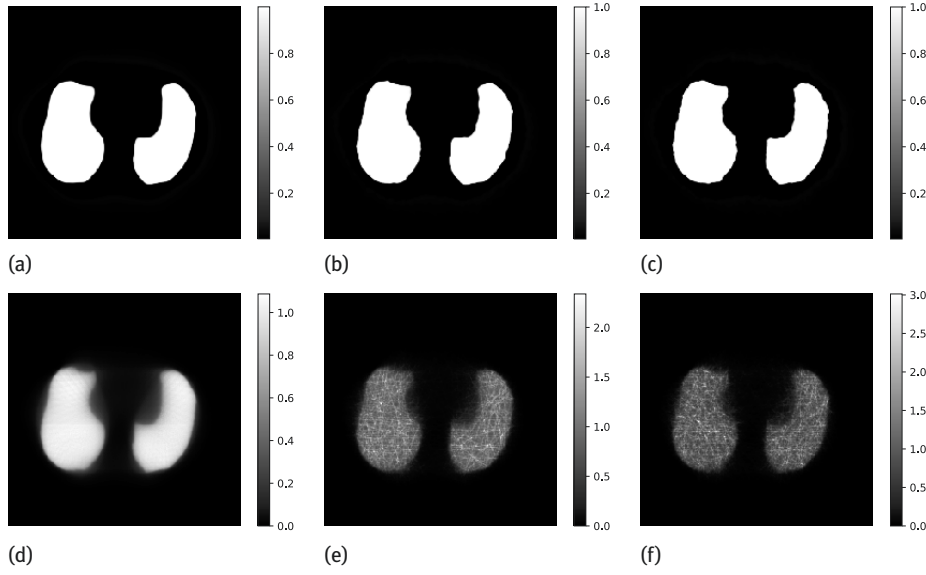


Figure 9: Lung perfusion imaging: solutions by $\mathcal{A}_\theta^{\dagger(\text{mlem})}$. (a)–(c) Final solutions when applying $\mathcal{A}_\theta^{\dagger(\text{mlem})}$ to the sinograms with 0, 10 %, and 20 % pepper and salt noises; (d)–(f) intermediate outputs of ML-EM in $\mathcal{A}_\theta^{\dagger(\text{mlem})}$.

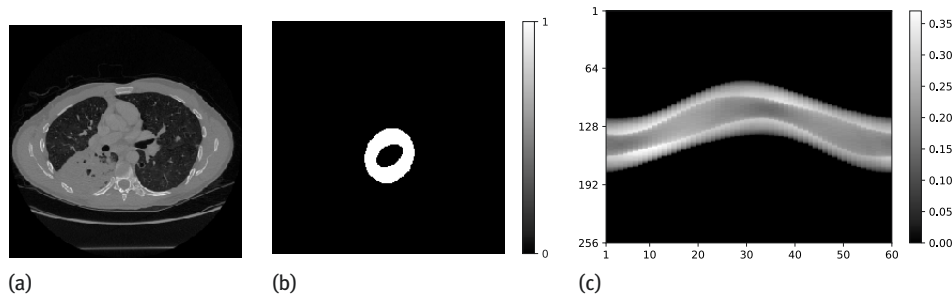


Figure 10: Myocardial perfusion imaging: setup of problem. (a) Attenuation map; (b) exact source image; (c) clean data sinogram without noises.

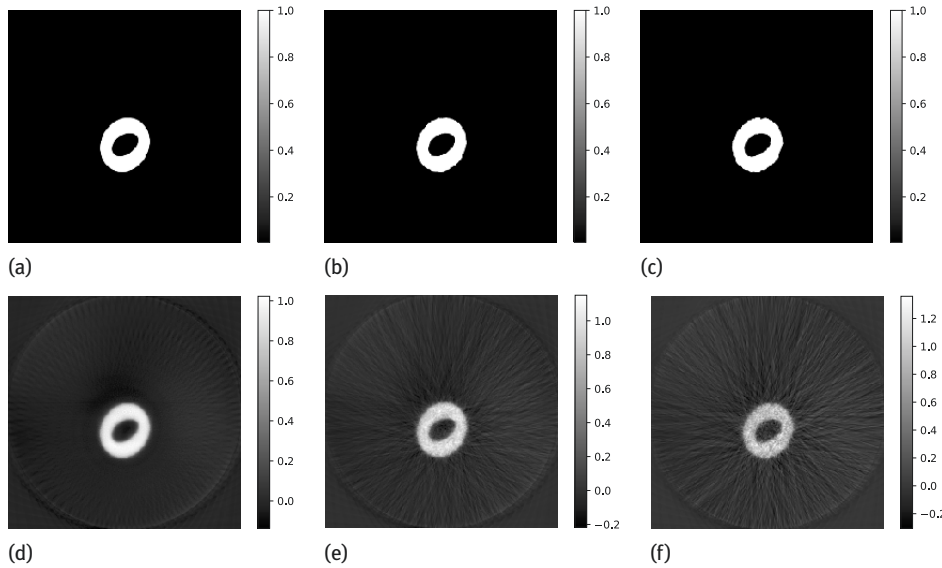


Figure 11: Myocardial perfusion imaging: solutions by $\mathcal{A}_\theta^{\dagger(\text{art})}$. (a)–(c) Final solutions when applying $\mathcal{A}_\theta^{\dagger(\text{art})}$ to the sinograms with 0, 10 %, and 20 % Gaussian noises; (d)–(f) intermediate outputs of ART in $\mathcal{A}_\theta^{\dagger(\text{art})}$.

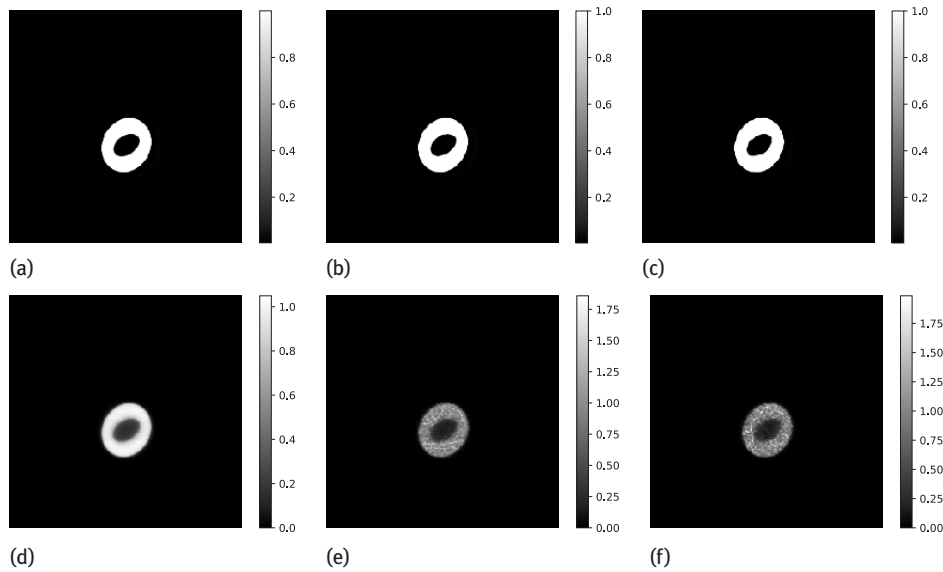


Figure 12: Myocardial perfusion imaging: solutions by $\mathcal{A}_\Theta^{\dagger(\text{mlem})}$. (a)–(c) Final solutions when applying $\mathcal{A}_\Theta^{\dagger(\text{mlem})}$ to the sinograms with 0, 10 %, and 20 % pepper and salt noises; (d)–(f) intermediate outputs of ML-EM in $\mathcal{A}_\Theta^{\dagger(\text{mlem})}$.

Pepper and salt noises	MSE		PNSR	
	Final solution	Intermediate output	Final solution	Intermediate output
0	0.000341	0.001652	35.20	27.35
10 %	0.000600	0.002380	32.51	20.84
20 %	0.000817	0.003155	31.08	20.55

Table 4: Myocardial perfusion imaging: MSE and PNSR of the final solutions and the intermediate outputs of $\mathcal{A}_\Theta^{\dagger(\text{mlem})}$.

5 Conclusion and future work

We have proposed a data and knowledge driven approach for SPECT that combines classical iterative algorithms with convolutional neural networks. The classical iterative algorithms, such as ART and ML-EM, are employed to include the model knowledge of SPECT, and the modified U-net is used to explore additional features of the source images and the sinogram data. We have developed an ART U-net and an ML-EM U-net, correspondingly. The proposed networks are trained according to supervised learning, where the mini-batch technique is used in the optimization process. We have applied the trained networks to the problems of simulated lung perfusion imaging and simulated myocardial perfusion imaging, respectively. Numerical examples demonstrate that the proposed networks are capable of reconstructing clean source functions even though a large amount of noise is introduced to the sinogram data.

In this work, we have assumed that the attenuation function is known so that we only reconstruct the source function in SPECT. The framework of the proposed networks, however, can be generalized to solve the identification problem of SPECT, where both the attenuation and source functions are to be determined. Firstly, the combined network will employ an iterative algorithm in the model-driven part so that it has the potential to handle the identification problem if an appropriate iterative algorithm is chosen.

Our work demonstrates that the performance of the combined network does not heavily rely on the selection of classical iterative algorithms since both ART and ML-EM lead to successful networks; therefore, it gives large flexibilities to the design of networks within the proposed framework. Secondly, our numerical examples illustrate that the combined networks are capable of reconstructing source images with piecewise constant structures. In the identification problem of SPECT, a piecewise constant source function is helpful

to alleviate ill-posedness and achieve uniqueness of the solution. In short, the proposed approach has the potential to solve the identification problem of SPECT, and we will study it in a future work.

Funding: Wenbin Li is supported by NSFC (grant no. 41804096), Natural Science Foundation of Guangdong Province (grant no. 2018A030313341), and Natural Science Foundation of Shenzhen (grant no. JCYJ2019 0806144005645). Qian's research is partially supported by NSF.

References

- [1] J. Adler and O. Öktem, Solving ill-posed inverse problems using iterative deep neural networks, *Inverse Problems* **33** (2017), no. 12, Article ID 124007.
- [2] J. Adler and O. Öktem, Learned primal-dual reconstruction, *IEEE Trans. Med. Imag.* **37** (2018), no. 6, 1322–1332.
- [3] E. V. Arbutov, A. L. Bukhgeim and S. G. Kazantsev, Two-dimensional tomography problems and the theory of A -analytic functions, *Siberian Adv. Math.* **8** (1998), no. 4, 1–20.
- [4] A. Beck and M. Teboulle, A fast iterative shrinkage-thresholding algorithm for linear inverse problems, *SIAM J. Imaging Sci.* **2** (2009), no. 1, 183–202.
- [5] J. Brownlee, A gentle introduction to mini-batch gradient descent and how to configure batch size, preprint (2017), <https://machinelearningmastery.com/gentle-introduction-mini-batch-gradient-descent-configure-batch-size/>.
- [6] L.-C. Chen, G. Papandreou, I. Kokkinos, K. Murphy and A. L. Yuille, Semantic image segmentation with deep convolutional nets and fully connected crfs, preprint (2014), <https://arxiv.org/abs/1412.7062>.
- [7] M. M. A. Dietze, W. Branderhorst, B. Kunnen, M. A. Viergever and H. W. A. M. de Jong, Accelerated SPECT image reconstruction with FBP and an image enhancement convolutional neural network, *EJNMMI Phys.* **6** (2019), Paper No. 14.
- [8] D. V. Finch, The attenuated x-ray transform: recent developments, in: *Inside Out: Inverse Problems and Applications*, Math. Sci. Res. Inst. Publ. 47, Cambridge University, Cambridge (2003), 47–66.
- [9] R. Gordon, R. Bender and G. T. Herman, Algebraic reconstruction techniques (ART) for three-dimensional electron microscopy and X-ray photography, *J. Theoret. Biol.* **29** (1970), no. 3, 471–481.
- [10] K. He, X. Zhang, S. Ren and J. Sun, Deep residual learning for image recognition, in: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, IEEE Press, Piscataway (2016), 770–778.
- [11] H. M. Hudson and R. S. Larkin, Accelerated image reconstruction using ordered subsets of projection data, *IEEE Trans. Med. Imag.* **13** (1994), no. 4, 601–609.
- [12] S. Ioffe and C. Szegedy, Batch normalization: Accelerating deep network training by reducing internal covariate shift, preprint (2015), <https://arxiv.org/abs/1502.03167>.
- [13] K. H. Jin, M. T. McCann, E. Froustey and M. Unser, Deep convolutional neural network for inverse problems in imaging, *IEEE Trans. Image Process.* **26** (2017), no. 9, 4509–4522.
- [14] D. P. Kingma and J. Ba, Adam: A method for stochastic optimization, preprint (2014), <https://arxiv.org/abs/1412.6980>.
- [15] A. Krizhevsky, I. Sutskever and G. E. Hinton, Imagenet classification with deep convolutional neural networks, in: *Advances in Neural Information Processing Systems*, Springer, New York (2012), 1097–1105.
- [16] D. S. Lalush and B. M. W. Tsui, A generalized gibbs prior for maximum a posteriori reconstruction in SPECT, *Phys. Med. Biol.* **38** (1993), no. 6, Paper No. 729.
- [17] S. Luo, J. Qian and P. Stefanov, Adjoint state method for the identification problem in SPECT: Recovery of both the source and the attenuation in the attenuated X-ray transform, *SIAM J. Imaging Sci.* **7** (2014), no. 2, 696–715.
- [18] R. G. Novikov, An inversion formula for the attenuated X-ray transformation, *Ark. Mat.* **40** (2002), no. 1, 145–167.
- [19] O. Ronneberger, P. Fischer and T. Brox, U-net: Convolutional networks for biomedical image segmentation, in: *International Conference on Medical Image Computing and Computer-Assisted Intervention*, Springer, Cham (2015), 234–241.
- [20] O. Russakovsky, J. Deng, H. Su, J. Krause, S. Satheesh, S. Ma, Z. Huang, A. Karpathy, A. Khosla and M. Bernstein, Imagenet large scale visual recognition challenge, *Int. J. Comput. Vis.* **115** (2015), no. 3, 211–252.
- [21] L. A. Shepp and Y. Vardi, Maximum likelihood reconstruction for emission tomography, *IEEE Trans. Med. Imag.* **1** (1982), no. 2, 113–122.
- [22] D. Strong and T. Chan, Edge-preserving and scale-dependent properties of total variation regularization, *Inverse Problems* **19** (2003), no. 6, S165–S187.
- [23] B. Wahlberg, S. Boyd, M. Annergren and Y. Wang, An ADMM algorithm for a class of total variation regularized estimation problems, *IFAC Proc.* **45** (2012), no. 16, 83–88.
- [24] X. Zhang, M. Burger and S. Osher, A unified primal-dual algorithm framework based on Bregman iteration, *J. Sci. Comput.* **46** (2011), no. 1, 20–46.