PAPER

Dynamic low-count PET image reconstruction using spatio-temporal primal dual network

IOP Publishing Phys. Med. Biol. 68 (2023) 135015 https://doi.org/10.1088/1361-6560/acde3e

Physics in Medicine & Biology

IPEM Institute of Physics and Engineering in Medicine

RECEIVED
20 March 2023

REVISED
12 May 2023

ACCEPTED FOR PUBLICATION
13 June 2023

PUBLISHED
30 June 2023

PAPER

Dynamic low-count PET image reconstruction using spatio-temporal primal dual network

Rui Hu[1], **Jianan Cui**[2], **Chenxu Li**[1], **Chengjin Yu**[1], **Yunmei Chen**[3] and **Huafeng Liu**[1,*]

[1] State Key Laboratory of Modern Optical Instrumentation, Department of Optical Engineering, Zhejiang University, Hangzhou 310027, People's Republic of China
[2] Institute of Information Processing and Automation, College of Information Engineering, Zhejiang University of Technology, Hangzhou 310001, People's Republic of China
[3] Department of Mathematics, University of Florida, Gainesville, FL 32611 United States of America
[*] Author to whom any correspondence should be addressed.

E-mail: rickhu@zju.edu.cn, jianancui@zjut.edu.cn, 22060538@zju.edu.cn, Chengjin.yu@gmail.com, yun@math.ufl.edu and liuhf@zju.edu.cn

Keywords: image reconstruction, positron emission tomography, model-based deep learning, low count, spatio-temporal correlation

Abstract

Objective. Dynamic positron emission tomography (PET) imaging, which can provide information on dynamic changes in physiological metabolism, is now widely used in clinical diagnosis and cancer treatment. However, the reconstruction from dynamic data is extremely challenging due to the limited counts received in individual frame, especially in ultra short frames. Recently, the unrolled model-based deep learning methods have shown inspiring results for low-count PET image reconstruction with good interpretability. Nevertheless, the existing model-based deep learning methods mainly focus on the spatial correlations while ignore the temporal domain. *Approach.* In this paper, inspired by the learned primal dual (LPD) algorithm, we propose the spatio-temporal primal dual network (STPDnet) for dynamic low-count PET image reconstruction. Both spatial and temporal correlations are encoded by 3D convolution operators. The physical projection of PET is embedded in the iterative learning process of the network, which provides the physical constraints and enhances interpretability. *Main results.* The experiments of both simulation data and real rat scan data have shown that the proposed method can achieve substantial noise reduction in both temporal and spatial domains and outperform the maximum likelihood expectation maximization, spatio-temporal kernel method, LPD and FBPnet. *Significance.* Experimental results show STPDnet better reconstruction performance in the low count situation, which makes the proposed method particularly suitable in whole-body dynamic imaging and parametric PET imaging that require extreme short frames and usually suffer from high level of noise.

1. Introduction

With the ability of quantifying biochemical and physiological parameters *in vivo*, dynamic positron emission tomography (PET) imaging plays an indispensable role in tumor detection (Kostakoglu *et al* 2003), characterization of heart diseases (Machac 2005) and drug development (Boss *et al* 2008). However, image reconstruction from dynamic PET data is challenging due to the ill-posedness of this inverse problem (Yu *et al* 2012, Liu *et al* 2015) and the low counting statistics of dynamic PET data, especially in the early-time frame reconstruction (Wernick *et al* 1999, Cui *et al* 2019a, 2019b).

Traditional reconstruction methods generally act on individual time frame data, include analytic methods, iterative methods and penalized log-likelihood methods. The images obtained by the analytical reconstruction methods such as filtered back-projection (FBP) (Brooks and Chiro 1976) usually have undesirable artifacts. Iterative methods such as maximum likelihood expectation maximization (MLEM) algorithm (Shepp and Vardi 1982) or its variant, the ordered-subsets accelerate version (Hudson and Larkin 1994) tend to fit the noise

characteristics in PET data well, while the results are limited by the accuracy of the projection model and the noise is increasing after a certain number of iterations. Penalized log-likelihood methods (Fessler 1994, Wang and Qi 2012, Chen *et al* 2015, Xie *et al* 2020) further improve the image quality by introducing the regularization term in the iteration framework. However, the design of the regularization is an open question and the hyper-parameters are often hand-crafted. These methods reconstructing from single time frame data does not take advantage of the time dependence in dynamic PET data.

By incorporating the temporal prior, spatio-temporal kernel methods (KEM-ST) (Wang and Qi 2014, Wang 2018) and other spatio-temporally constrained strategies (Chen *et al* 2015, Zhang and Liu 2019) improves the reconstruction. Nevertheless, for KEM-ST, global temporal kernels may not match the local temporal characteristics in some atypical regions and the large temporal window size possibly result in over-smoothing. Besides, the ultra short time frames are often required in clinical dynamic PET imaging, while these methods may not perform well in such low counting situation.

As deep learning methods show extraordinary capabilites in medical imaging, many deep learning-based PET image reconstruction methods have been proposed (Gong *et al* 2019, Reader *et al* 2020), including direct learning methods (Zhu *et al* 2018), post-denoising methods (Cui *et al* 2019c, Hashimoto *et al* 2019, Zhou *et al* 2020, Hashimoto *et al* 2021, Onishi *et al* 2021) and deep unrolling methods (Gong *et al* 2019, Lim *et al* 2020, Mehranian and Reader 2020, Hu and Liu 2022). Direct learning method are easy to implement, such as DeepPET (Häggström *et al* 2019, Li *et al* 2022). However, it always requires a large number of training data pairs and lacks interpretability and generalization ability (Whiteley *et al* 2020). Post-denoising methods are the most common way to apply deep learning techniques in PET. However, post-processing cannot save the lengthy reconstruction time and its results are significantly affected by the pre-reconstruction algorithm (Cheng *et al* 2021). As an emerging technology, deep image prior (DIP) is widely used in dynamic PET denoising. Cui *et al* (2019c) used the DIP for unsupervised dynamic PET denoising. Hashimoto *et al* (2019) utilized CNN and DIP for dynamic PET image denoising. However, these DIP based methods are easy to overfit, which limits the accuracy of parameter estimation. Deep unrolling methods combine the physical characteristics of PET imaging model with the powerful representation ability of deep neural network and show inspiring reconstruction results. Mehranian *et al* (Mehranian and Reader 2020) unrolled a forward backward splitting algorithm with CNN for PET image reconstruction. Lim *et al* (2020) unrolled the block coordinate descent algorithm with Unet. However, most of these methods focus only on the reconstruction of image domain and do not catch the temporal characteristics of the dynamic PET data. FBPnet (Wang and Liu 2020) is one of the few deep learning-based methods that takes into account the temporal relationships of dynamic PET data. It combines the traditional FBP algorithm and a denoising neural network and has good generalization ability. Nevertheless, this methoddoes not do iterative refining, just like the FBP algorithm.

In this paper, we propose an unrolled deep learning method, the spatio-temporal primal dual network (STPDnet), for dynamic low count PET image reconstruction. As one of the model-based deep learn method, the proposed STPDnet has good interpretability, as well as excellent learning ability. The main contributions of this paper include (1) introducing the STPDnet (Spatio-Temporal Primal Dual network), the first unrolled neural network for dynamic low count PET image reconstruction; (2) using 3D spatio-temporal convolution operators to simultaneously model the spatial and temporal correlations in dynamic measured sinogram. Part of this work was previously accepted to the 2023 IEEE International Symposium on Biomedical Imaging (ISBI) (Hu *et al* 2023). This work has been substantially extended with more detailed studies on the simulation data as well as experimental evaluations of clinical rat data. Additionally, we present a comprehensive discussion of the experimental results obtained from the proposed method, along with an analysis of its strengths and limitations. Furthermore, a detailed literature review of related works is also incorporated in this paper.

The rest of this paper is organized as follows. Section 2 introduces the network structure of proposed STPDnet and details of the reconstruction process. Section 3 describes the simulations, data set and the evaluation methods. The experimental results are shown in section 4, followed by further discussion of the results in section 5. Section 6 gives the final conclusions.

2. Method

2.1. Dynamic PET image reconstruction

Dynamic PET imaging enable real-time tracking of radiotracers, providing valuable insights into organ function and metabolism by extracting and analyzing the kinetic parameters of the reconstructed dynamic PET images. A sequence of dynamic PET images is reconstructed from a time-dependent series of sinograms, providing a comprehensive representation of the temporal evolution of the radiotracer distribution within the subject.

For dynamic PET imaging, the expectation of the measurement data $\bar{\mathbf{y}} = \{\bar{y}_{i,t}\} \in \mathbb{R}^{I\times 1}$ in an individual frame t can be modeled by a linear transformation as:

$$\bar{\mathbf{y}} = G \cdot \mathbf{x} + \mathbf{r}, \tag{1}$$

where $\mathbf{x} = \{x_{j,t}\} \in \mathbb{R}^{J\times 1}$ is the unknown image, $G \in \mathbb{R}^{I\times J}$ is the system response matrix with j representing the index of pixels and i representing the index of detectors. $G_{i,j}$ denotes the probability that the ith detector received the photon emitted from the jth voxel (Qi *et al* 1998). I is the total number of detector pairs and J is the total number of voxels. $\boldsymbol{r} \in \mathbb{R}^{I\times 1}$ is the expectation of dynamic random and scattered events.

Dynamic PET image reconstruction is an ill-posed problem, which can be solved by minimizing a regularized objective function:

$$\min_{\mathbf{x}\in X} - L(\mathbf{y}|\mathbf{x}) + \lambda R(\mathbf{x}) \tag{2}$$

$$L(\mathbf{y}|\mathbf{x}) = \sum_{i=1}^{I}\sum_{t=1}^{T} y_{i,t}\log \bar{y}_{i,t} - \bar{y}_{i,t}, \tag{3}$$

where X denotes the reconstructed image space. $L(\mathbf{y}|\mathbf{x})$ is the negative Poisson log-likelihood, which also called the data fidelity term. T is the total number of time frames. $R(\mathbf{x})$ is the regularization term controlled by the parameter λ.

2.2. Primal dual hybrid gradient (PDHG) algorithm for PET

The dynamic PET image reconstruction problem can be reformulate to the saddle point problem (Ehrhardt *et al* 2019):

$$\min_{\mathbf{x}\geqslant 0}\sup_{\mathbf{y}} < G\mathbf{x} + \mathbf{r}, \mathbf{y} > -L^*(\mathbf{y}) + \lambda R(\mathbf{x}), \tag{4}$$

where L^* is the convex conjugate of the log-likelihood L (Bauschke *et al* 2011) and $< \cdot, \cdot >$ denotes the inner product with $<\mathbf{A}, \mathbf{B}> = Tr(\mathbf{A}^T, \mathbf{B})$.

The saddle point problem equation (4) can be solved by PDHG algorithm (Chambolle and Pock 2011) with an introduced dual variable $\mathbf{h} \in Y$. $\mathbf{Y}$ is the measurement sinogram space. The scheme of PDHG algorithm for dynamic PET image reconstruction is shown in algorithm 1. The iteration process is divided into three steps. In the first step, the sum of the measured sinogram and the forward projection of the estimated image is calculated and then evaluated by the proximal operator. The proximal operator is defined by:

$$\text{prox}_{mL^*} = \underset{\tilde{\mathbf{x}}\in X}{\text{argmin}}\{||\tilde{\mathbf{x}} - \mathbf{x}||^2 + L^*(\mathbf{x})\}. \tag{5}$$

After the dual variable $\mathbf{h}^0_{i,t}$ is updated (line 2), the back-projection is calculated. Similarly, the primal variable $\mathbf{x}$ is updated through the proximal operator (line 3).

Algorithm 1. Primal-dual hybrid gradient (PDHG) to solve the dynamic PET image reconstruction.

Input: image initialization $\mathbf{x}^0_{j,t}$, dual variable initialization $\mathbf{h}^0_{i,t}$, measured sinogram $\mathbf{y}_{i,t}$, step size $m = \{m_{i,t}\}, n = \{n_{j,t}\}, p = \{p_{i,t}\}$
1: **for** $i \in [1,\ldots]$ **do**
2: $\quad \mathbf{h}^k_{i,t} \leftarrow \text{prox}_{mL^*}(\mathbf{h}^{k-1}_{i,t} + m\mathbf{G}\cdot \bar{\mathbf{x}}^{k-1}_{j,t})$
3: $\quad \mathbf{x}^k_{j,t} \leftarrow \text{prox}_{nR}(\mathbf{x}^{k-1}_{j,t} - n\mathbf{G}^*\cdot \mathbf{h}^k_{i,t})$
4: $\quad \bar{\mathbf{x}}^k_{j,t} = \mathbf{x}^k_{j,t} + p(\mathbf{x}^k_{j,t} - \mathbf{x}^{k-1}_{j,t})$
5: **end for**

2.3. Spatio-temporal convolutional primal dual network

Previously, the learned primal dual (LPD) (Adler and Öktem 2018) used two networks to represent the proximal operators. The prior information and the update process can be learned from the training data. Although it has achieved good results on CT, CS-MRI (Wang *et al* 2019) and static PET (Guazzo and Colarieti-Tosti 2021), it is not suitable for dynamic PET data with low count properties and strong temporal correlations. The main reasons are as follows: (1) by solely employing convolutional and activation function to fit the proximal operator, gradient explosion issues may arise, making it challenging for the network to converge to a satisfactory result, especially in the case of low count PET data. (2) The temporal correlations of measurement data is not considered. Especially in dynamic PET data, there is a strong dependence of different time frames.

Based on these observations, we proposed the STPDnet for low count dynamic PET image reconstruction. Firstly, we replace the two proximal operators in the PDHG algorithm with the primal net P and the dual net D based on spatio-temporal 3D convolution. 3D convolution is carried out in both spatial and temporal domains to simultaneously extract the spatio-temporal features. The primal net is used to learn the spatio-temporal correlations in the reconstructed image domain and update the primal variable $\mathbf{x}_{j,t}$. The dual net is used in the measurement sinogram domain to update the dual variable $\mathbf{h}_{i,t}$. Thus, the dual domain alternate iteration is

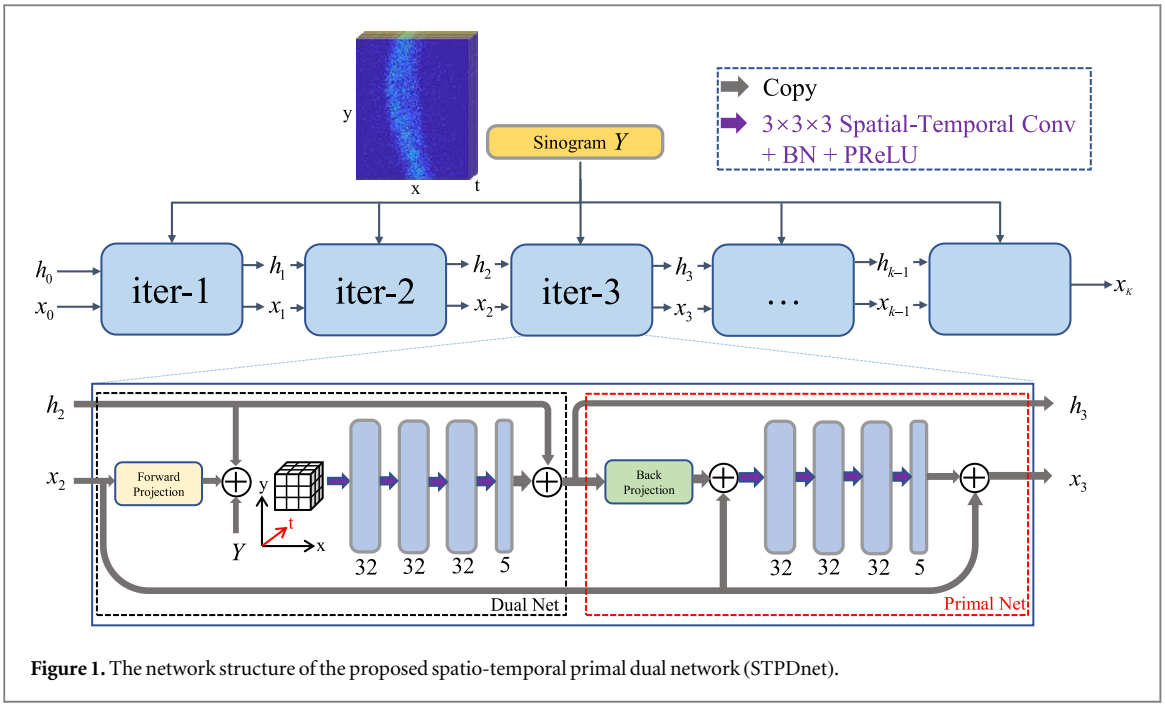

Figure 1. The network structure of the proposed spatio-temporal primal dual network (STPDnet).

formed. Meanwhile, we let the network learn the relationship between the input items of the proximal operator:

$$\mathbf{h}_{i,t}^{k} = D(\mathbf{h}_{i,t}^{k-1}, \mathbf{G} \cdot \mathbf{x}_{j,t}^{k-1}, \mathbf{y}) \tag{6}$$

$$\mathbf{x}_{j,t}^{k} = P(\mathbf{x}_{j,t}^{k-1}, \mathbf{G}^{*} \cdot \mathbf{h}_{i,t}^{k}), \tag{7}$$

where, $D(\cdot) = P(\cdot) = \mathrm{BN}(w_l^{3d} \cdot \cdot\cdot\sigma(\mathrm{BN}(w_2^{3d} * \sigma(\mathrm{BN}(w_1^{3d}(\cdot))))))$ denotes the structure of primal net and dual net. $\{w_q^{3d}\}_{q=1}^{l}$ is the convolution weights consisting of kernels with spatio-temproal kernel size $(3 \times 3 \times 3)$, BN denotes the batch normalization layer, σ is the PReLU activation function and $*$ is the convolution operation.

Compared with LPD, the 3D spatio-temporal convolution is adopted and the batch normalization (BN) is introduced behind every convolution operator to enable the network to learn more efficiently. The batch normalization layer makes the network converges faster. In order to reduce the effect of low accuracy of projection operator on reconstruction results, we add a convolution layer after each projection to learn the gap between the simulated PET projection and the real world measurement. Techniques like BN and skip connection allow the network with better generalization ability and make it particularly suitable for the dynamic PET image reconstruction in low count situation. The extended primal memory and dual memory are also introduced in the STPDnet, which are obtained by appending a set of 'lifting variables' to the original primal and dual variables. These lifting variables can make the optimization problem more amenable to specific operations, such as convolution. The overall algorithm flowchart is presented in algorithm 2. The proposed network contains K iterations, each iteration corresponds to one block as shown in figure 1.

Algorithm 2. Algorithm for dynamic PET reconstruction with spatio-temporal convolutional primal dual net-work.

Input: image initialization $\mathbf{x}_{j,t}^{0} \in^{(n_{\text{primal}})} \mathbf{X}$, dual variable initialization $\mathbf{h}_{i,t}^{0} \in^{(n_{\text{dual}})} \mathbf{Y}$, maximum number of iteration blocks K, measured Sinogram $\mathbf{y}$

1: **for** $k \in [1, K]$ **do**
2: $\quad \mathbf{h}_{i,t}^{k} \leftarrow D(\mathbf{h}_{i,t}^{k-1}, \mathbf{G} \cdot^{(2)} \mathbf{x}_{j,t}^{k-1}, \mathbf{y})$
3: $\quad \mathbf{x}_{j,t}^{k} \leftarrow P(\mathbf{x}_{j,t}^{k-1}, \mathbf{G}^{*} \cdot^{(1)} \mathbf{h}_{i,t}^{k})$
4: **end for**
5: **return** $^{(1)}\mathbf{x}^{K}$;

2.4. Implementation details

The proposed network was implemented using Pytorch 1.7 on a NVIDIA TITAN-X of 24 GB graphics card memory. The number of iteration blocks (K) was 8. The dual variable $\mathbf{h}_{i,t}^{0}$ and image $\mathbf{x}_{j,t}^{0}$ were both initialized with values of zero. During training, the dynamic low count sinogram was fed into the network and the mean

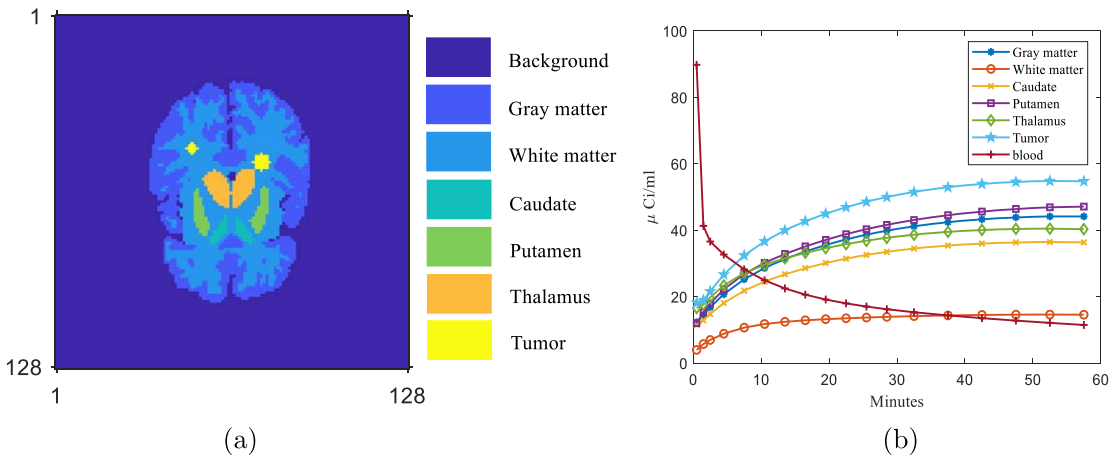

Figure 2. The Zubal phantom and time activity curves (TACs) used in the simulation study. (a) The 35th slice of the 3D Zubal phantom; (b) examples of the regional time activity curves.

Table 1. The mean values of the simulated K parameters. V denotes blood volume ratio. The standard deviation (STD) of each value is the 10% of its mean.

Tissue	K_1	k_2	k_3	k_4	V
Gray matter	0.100	0.140	0.170	0.013	0.103
White matter	0.050	0.110	0.050	0.006	0.026
Caudate	0.120	0.170	0.190	0.016	0.101
Putamen	0.130	0.160	0.170	0.010	0.092
Thalamus	0.130	0.160	0.140	0.012	0.152
Tumor	0.110	0.100	0.150	0.015	0.173

square error (MSE) loss was calculated between the network outputs and the label images. The learnable parameters, include convolution kernels, biases, batch normalization parameters were optimized by Adam optimizer with $\beta_1 = 0.9$ and $\beta_2 = 0.999$. The learning rate was 8e-4 and decayed by a factor of 0.99 after each epoch. The number of extended memory of the primal variable (n_{primal}) and dual variable(n_{dual}) were 3. The batchsize was 2.

3. Experimental setup

3.1. Data set

3.1.1. Simulation data

The 128 × 128 × 40 3D Zubal brain phantom (Zubal *et al* 1994) was used in the simulation study and two tumors with different size were added. In the simulation, the gray matter to white matter ratio of each phantom was modeled as a Gaussian variable with the variation of 0.1. The phantom contains 6 regions of interests (ROIs) as shown in figure 2(a). The scanning schedule consisted of 18 time frames over 60 min: 3 × 60 s, 9 × 180 s, 6 × 300 s.

A three compartmental model with Feng's input function (Feng *et al* 1993) was adopted to simulate dynamic ^{18}F-FDG scans. Regional TACs as shown in figure 2(b) were assigned to different regions to generate true tracer radioactivity images. To simulate the population difference, k parameters were modeled as a Gaussian variable with standard deviations of 0.1. The mean values of these k parameters were presented in table 1.

The system matrix used in the data projection was computed by using Michigan Image Reconstruction Toolbox (Noh *et al* 2009) with a strip-integral model for both simulation data and real rat data. The true tracer radioactivity images were first forward projected to generate noise-free sinogram. The number of detector bins and projection angles were 128 and 160. Poisson noise were introduced to the noise free sinogram with 20% of the counts. The total count of one simulated brain sample was approximately 2.4×10^5 to simulate low count situation. For a single slice sinogram, frame 1 only had about 5k events, whereas frame 18 had about 17k events. In total, 40 4D training pairs with size of 128 × 128 × 40 × 18 were generated, where 33 were randomly selected as training data, 5 as the test data, 2 as the validation data. As our network needs both temporal and spatial data as

input, each axial slice with 18 time frames (128 × 128 × 18) was used as one training sample, making total of 1600 data samples (1320 for training, 80 for validation, 200 for testing), the total image slices are 28800. The training label is the noise-free activity images. A total of 5 realizations were simulated and each was trained and tested independently for bias-variance analysis.

3.1.2. Rat data

Animal experiments were approved by the experimental animal welfare and ethics review committee, and were performed in compliance with local legal requirements. Thirteen rat with gliomas datasets of one hour FDG dynamic scan acquired on Siemens Micro-PET/CT Inveon scanner with 1 mCi dose injection were employed in this study. Data acquisition began right after the FDG injection. The scanning schedule consisted of 18 time frames over 60 min: 3 × 60 s, 9 × 180 s, 6 × 300 s. Only the segments 0 of the michelograms were used as training and test sinogram (160 views, 128 bins) for low count simulation. For a single slice sinogram, frame 1 has 5k events, whereas frame 18 has 20k events approximately. These low-count sinogram was taken as inputs, the reconstructed PET images with CT attenuation correction and full 3D counts were used as labels. The images reconstructed of one rat data sample with an array size of 128 × 128 × 120 × 18. The system matrix used in the reconstruction is the same as the simulation. 10 rats were selected randomly for training, 1 for validation and 2 for testing.

3.2. Data evaluation

Peak signal-to-noise ratio (PSNR) and the structural similarity (SSIM) were used for overall image quality evaluations of both simulation experiments and rat experiments. For quantitative comparison and bias variance analysis in simulation experiments, contrast recovery coefficient (CRC) versus the standard deviation (STD) curves and the bias versus the standard deviation curves were plotted. For the comparison of the accuracy of reconstruction TACs, the Euclidean distances between the reconstructed TACs and label TACs were calculated. Furthermore, the Patlak graphical method is used for quantitative parameter comparisons.

The PSNR is defined by:

$$\mathrm{PSNR} = 20 \cdot \log_{10}\left(\frac{q_{\max}}{\sqrt{\frac{1}{n}\sum_{i=1}^{n}(p_i - q_i)^2}}\right), \tag{8}$$

where p is the reconstructed image, q is the ground truth, n is the number of image pixels, $q_{\max}$ is the maximum value of the ground truth image.

The SSIM is defined by:

$$\mathrm{SSIM} = \frac{(2\mu_p\mu_q + c_1)(2\sigma_{pq} + c_2)}{({\mu_p}^2 + {\mu_q}^2 + c_1)({\sigma_p}^2 + {\sigma_q}^2 + c_2)} \tag{9}$$

where μ and σ denote the mean of the image and the variance of the image. $c1 = 0.01 \times \max(p)$ and $c2 = 0.03 \times \max(p)$.

The CRC is defined as:

$$\mathrm{CRC} = \frac{1}{R}\sum_{r=1}^{R}\frac{\left(\frac{\bar{a}_r}{\bar{b}_r} - 1\right)}{\left(\frac{a_{\mathrm{true}}}{b_{\mathrm{true}}} - 1\right)}. \tag{10}$$

Here, R denotes the number of realizations. $\bar{a}_r$ and $\bar{b}_r$ are the average values of selected ROIs and background regions in rth realization. a_{true} and b_{true} are the ground truth values of the target and the background region, respectively. In the simulation study, the number of realizations is 5. The two tumor regions were selected as target ROIs and 8 four-pixel-size-diameter spheres were drawn in whiter matter as background region. The background STD is calculated as:

$$\mathrm{STD} = \frac{1}{K_b}\sum_{k=1}^{K_b}\frac{\sqrt{\frac{1}{R-1}\sum_{r=1}^{R}(b_{r,k} - \bar{b}_k)^2}}{\bar{b}_k}, \tag{11}$$

where $K_b = 8$ is the total number of ROIs in the background region and $\bar{b}_k = (1/R)\sum_{r=1}^{R} b_{r,k}$ is the average of the kth ROI over R realizations.

The bias of the mean values in the whole target region is defined by:

$$\mathrm{Bias} = \frac{\bar{p} - q}{p}, \tag{12}$$

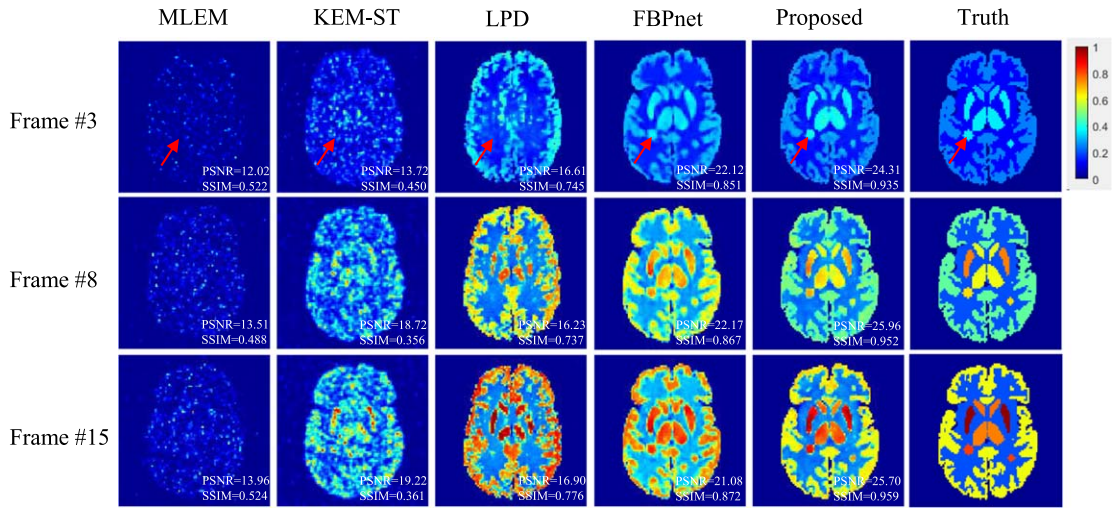

Figure 3. The ground truth images and reconstructed images of 35th slice by different reconstruction methods for the 3th frame (top row), 8th frame (middle row) and 15th frame (bottom row). From left to right: MLEM, KEM-ST, LPD, FBPnet, proposed STPDnet and ground truth.

where $\bar{p}$ is the average value of whole target region for all realizations, q is the ground truth value of the whole target region.

Under the Patlak graphical method, the tracer concentration at time t is represented by a weighted sum of the blood input function $C_P(t)$ and its integral after a sufficient lenght of time t^* as:

$$\boldsymbol{c}(t) = \boldsymbol{\kappa} \int_0^t C_p(\tau)d\tau + \boldsymbol{b}C_p(t),\ t > t^\star, \tag{13}$$

where $\boldsymbol{\kappa}$ represent the Patlak slope image, and $\boldsymbol{b}$ denote the Patlak intercept image. In practice, the Patlak slope image $\boldsymbol{x}$ is a crucial quantitative metric as it reflects the influx rate of the PET tracer.

3.3. Reference methods

Both simulation and rat noisy sinogram was reconstructed independently by five methods: the traditional MLEM (Shepp and Vardi 1982), spatio-temporal kernel method with the data-driven temporal kernel (KEM-ST) (Wang 2018), LPD (Adler and Öktem 2018), FBPnet (Wang and Liu 2020) and proposed STPDnet. The MLEM and KEM-ST were run 30 iterations, the iteration with the best PSNR and SSIM were chosen for comparison. The spatial and temporal kernel matrixes were constructed using the same approach as described in (Wang 2018). For LPD, FBPnet and STPDnet, MSE loss and and Adam optimizer were used for training. The total training epochs were 300, and the optimal epoch with the minimum validation loss were selected as the final reconstruction model. The network structure of LPD and FBPnet were set as same as the original paper. The training setting of three learning methods were same.

4. Results

4.1. Simulation results

Figure 3 shows the 35th slice axial view of the reconstructed images by five different methods and the ground truth images for the 3th frame, 8th frame and 15th frame, respectively. It can be observed that little structural information was recovered by MLEM, and only a few edge contours were visible in ultra-low count situation. KEM-ST achieved a significant PSNR increase as compared with MLEM, while the structure information was still missing, especially in the early-time frame. LPD revealed more cortex information and improved the SSIM than KEM-ST and MLEM. However, due to the lack of temporal information modeling, the reconstructed images of different frames by LPD showed significant differences. Besides, the tumor regions that the red arrow points to had barely been recovered. Figure 4 shows the mean image PSNR and mean image SSIM over 18 time frames of the five different methods. KEM-ST had higher PSNR than MLEM, but some structural similarities were lost. The three learning-based methods outperformed the MLEM and KEM-ST for all frames. It can be seen that the SSIM of FBPnet showed a significant drop in frame six. We checked out the results and found that the sixth frame images reconstructed by FBPnet appeared a significant over-smoothing. Compared with the LPD and FBPnet, the proposed STPDnet substantially improved PSNR and SSIM and showed stable results. Figure 5 shows the CRC versus STD curves at the tumor regions using different reconstruction methods. It can be seen

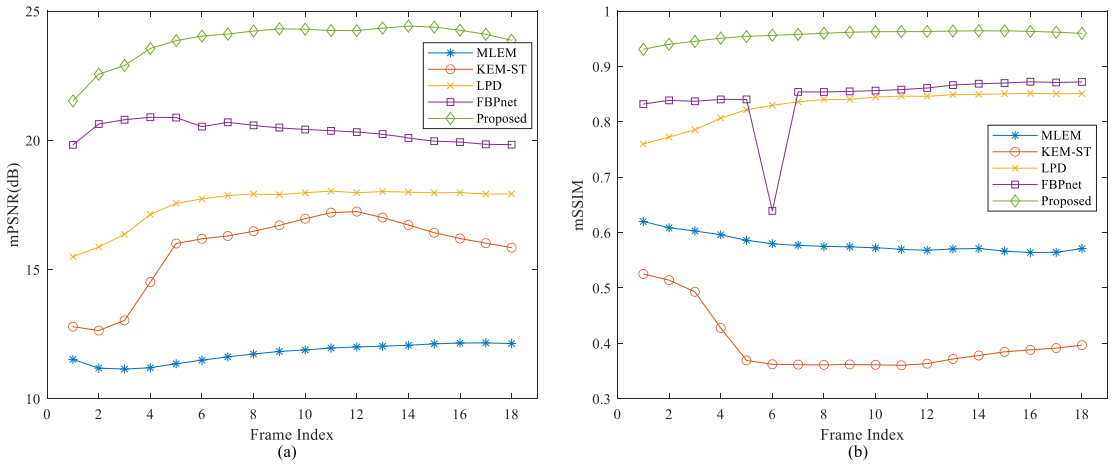

Figure 4. Plots of mean image PSNR and mean image SSIM of simulation test set over 18 time frames reconstructed by different methods. (a) The mean of PSNR of the test set; (b) the mean of SSIM of the test set.

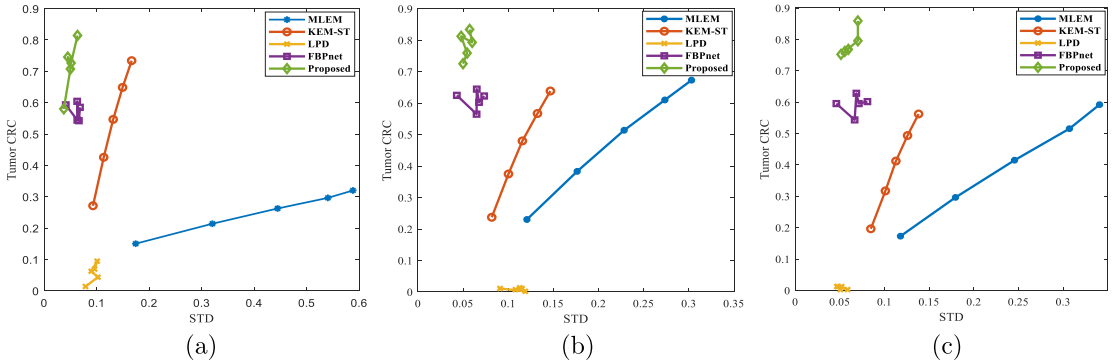

Figure 5. The CRC versus STD curves of tumor ROI by varying iteration numbers or training epochs for the (a) 3th frame, (b) 8th frame, (c) 15th frame. For the MLEM and KEM-ST, markers were plotted every 5 iterations; for LPD, FBPnet and proposed STPDnet, markers were plotted every 50 epochs.

that the proposed STPDnet achieved high CRC with lower STD than other four compared methods. The CRC of LPD was very low because LPD can hardly recover the tumor region with the poor learning ability and the lack of temporal modeling. Figure 6 shows the STD versus Bias curves of whole target regions. The proposed STPDnet had the lowest Bias in the same STD level than other four comparison methods. Both CRC versus STD curves and STD versus Bias curves demonstrated that the proposed STPDnet outperformed other methods in bias-variance trade-off. Figure 7 shows the TACs of the tumor region reconstructed by five different methods. The TAC reconstructed by MLEM fluctuated widely and showed extreme noise. KEM-ST achieved a good noise reduction in the temporal domain, while the accuracy needed further improvement. LPD also showed obvious noise due to the lack of temporal modeling. The TAC reconstructed by FBPnet was very smooth but the gap with the ground truth TAC was still large. The proposed STPDnet performed the best both in the temporal noise reduction and in the TAC accuracy. To measure the TAC accuracy of different reconstruction methods in the whole tumor region, The mean L2 distance between the reconstructed TAC and the ground truth TAC for each pixel in the tumor region was calculated as shown in table 2. As seen, the proposed STPDnet got the most accurate TAC compared with other methods, which shows that STPDnet has a good reconstruction ability in the temporal domain. Figure 8 shows the Patlak slopes reconstructed by different methods and the ground truth K_i image calculated by using compartmental models. It can be seen that the noise of the Patlak slopes obtained by MLEM and KEM is very severe. Under low count conditions, traditional methods are difficult to obtain accurate Patlak slopes. LPD has a obvious noise suppression, but the structural information is not well recovered, especially in the simulated tumor region. FBPnet has significant improvements in noise suppression and structure recovery, but the overall K_i value is high, and the tumor boundary is not clear. The proposed method performs the best among all comparison methods.

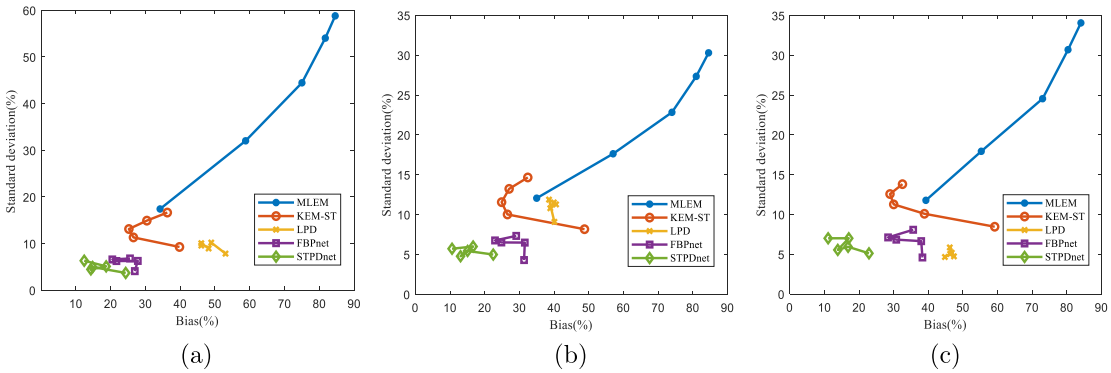

Figure 6. The STD versus Bias curves by varying iteration numbers or training epochs for the (a) 3th frame, (b) 8th frame, (c) 15th frame. For the MLEM and KEM-ST, markers were plotted every 5 iterations; for LPD, FBPnet and proposed STPDnet, markers were plotted every 50 epochs. Bias was calculated based on the whole target region.

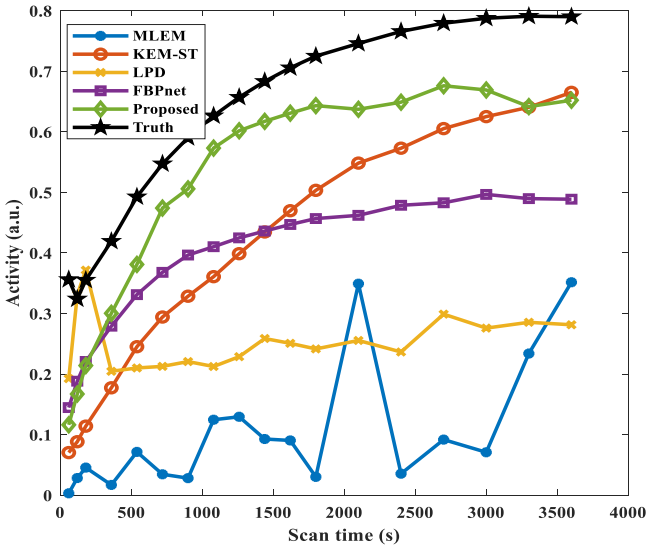

Figure 7. Time activity curves of tumor region reconstructed by different reconstruction methods.

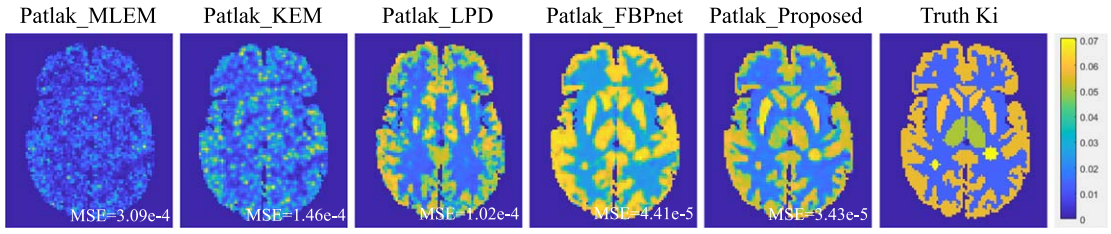

Figure 8. The reconstructed Patlak slopes of 35th slice of the test set using different methods.

Table 2. The L2 distance between the ground truth TACs and reconstructed TACs by different methods in the tumor region.

	MLEM	KEM-ST	LPD	FBPnet	Proposed
L2 distance (meanstd)	2.450.006	1.650.28	1.750.21	0.760.43	**0.67 ± 0.46**

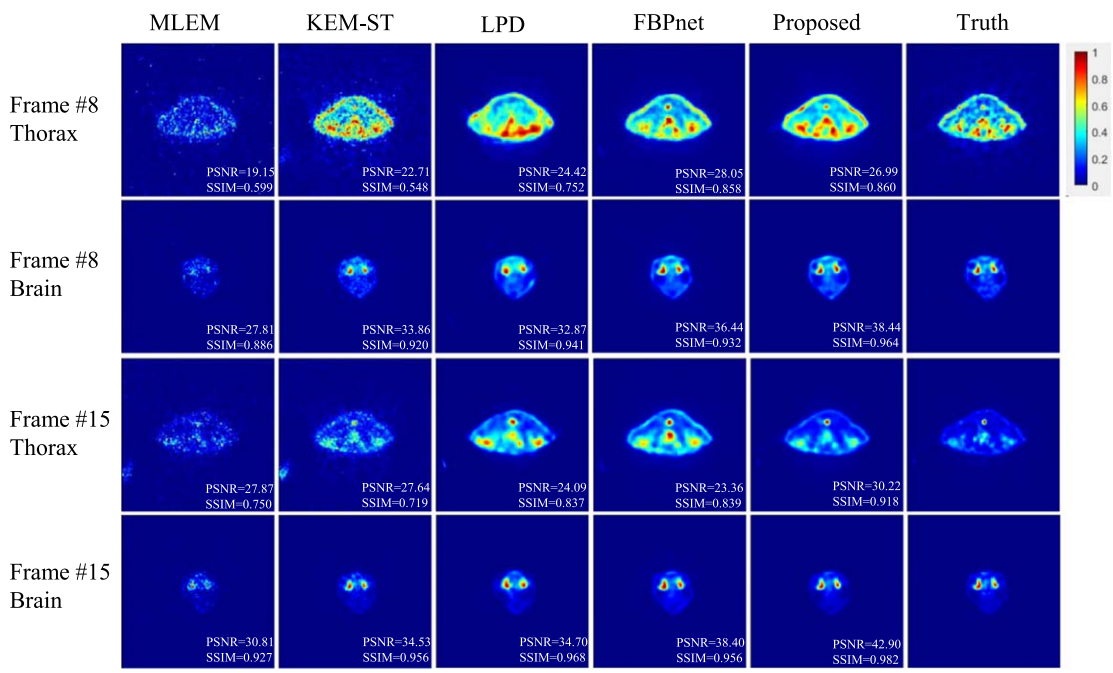

Figure 9. Reconstruction images of rat data by different methods. From top to bottom: thorax slice in frame 8, brain slice in Frame 8, thorax slice in frame 15, brain slice in frame 15.

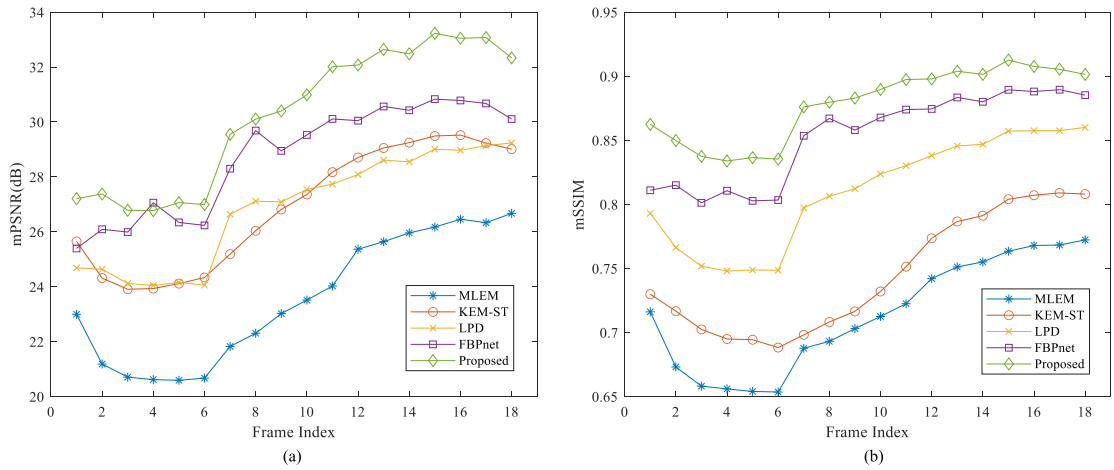

Figure 10. Plots of mean image PSNR and mean image SSIM of rat test set over 18 time frames reconstructed by different methods. (a) The mean of PSNR of the test set; (b) the mean of SSIM of the test set.

4.2. Rat data results

Figure 9 shows the reconstruction images of whole body data of rats by different methods in frame 8 and frame 15. It can be observed that traditional MLEM was very noisy, KEM-ST can distinguish among hot region of suspected tumor to some extent, but the noise of image was very serious. LPD showed relatively smooth images, but misjudged some hot regions of suspected tumors badly, especially in some early-time frames such as frame 8. The same misjudgement occurred in the reconstruction results of FBPnet. The proposed STPDnet achieved the best performance both in the overall contour of the image and in the detailed area of suspected tumor. Figure 10 shows the mean image PSNR and mean image SSIM of rat test set over 18 time frames achieved by the five different methods. As the scanning time increased, the counts were getting higher, the PSNR and SSIM showed a rising trend. In general, the mean image PSNR and mean image SSIM of rat test set by proposed STPDnet were higher than those by other comparison methods. Figure 11 shows the box plots of PSNR and SSIM for all frames, the proposed STPDnet achieved the highest PSNR and SSIM over all time frames. Figure 12 shows the J-shape tumor ROI in the rat brain and the box plot of the L2 distance between reconstructed TAC and full dose TAC in

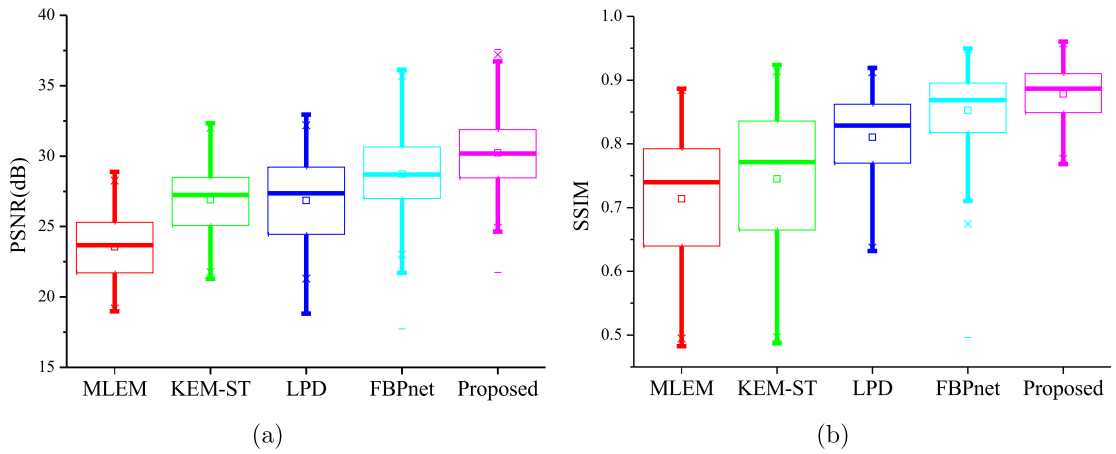

Figure 11. The box plots of PSNR and SSIM of rat test set for all frames. (a) The box plot of PSNR of the rat test set; (b) the box plot of SSIM of the rat test set.

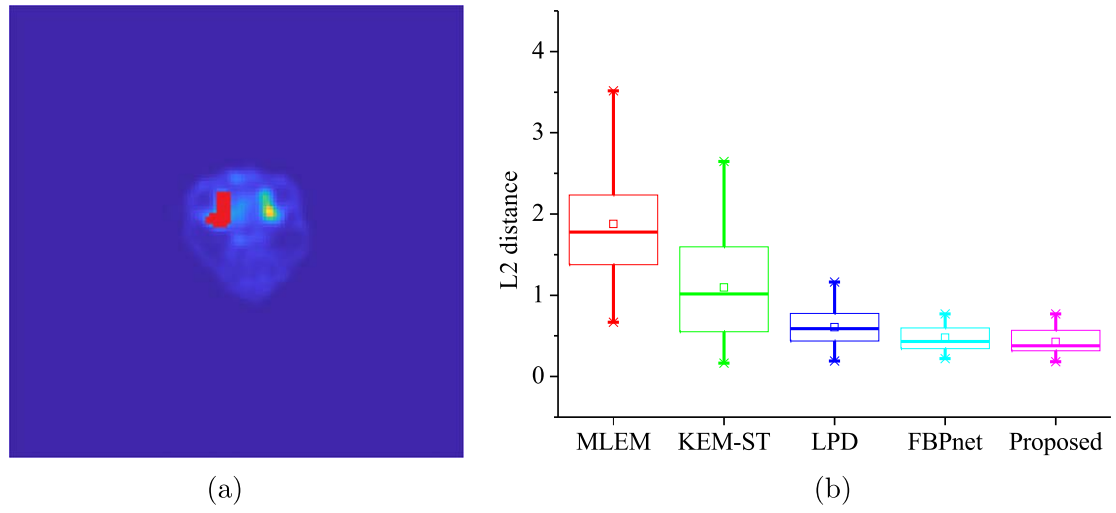

Figure 12. J-shape tumor ROI and the box plot of the L2 distance between reconstructed TAC and full dose TAC in the pixels of tumor ROI region. (a) The rat brain with marked J-shape tumor ROI. (b) The box plot of the L2 distance.

the pixels of tumor ROI region. It can be seen that the proposed method outperformed other comparison methods in terms of reconstructed TAC accuracy and stability in the temporal domain.

5. Discussion

In this study, we propose an unrolled deep learning method, STPDnet, for dynamic low count PET image reconstruction, which achieved better performance compared MLEM, KEM-ST, LPD and FBPnet in both simulation data and real rat data. In this section, we give some discussions and comments on the STPDnet.

5.1. The number of unrolled blocks

As one of the deep unrolled methods, the performance of STPDnet is highly related to the number of unrolled blocks. Within a certain range, increasing the number of unrolled blocks can enhance the network's learning and representation capabilities. However, it also leads to issues such as large memory usage and extended training times. Moreover, the number of unrolled blocks cannot be increased indefinitely, as this would result in severe over-fitting. So there are trade-offs between network effects and time or space consumption. We have trained and tested STPDnet under different numbers of the unrolled block, and quantified the reconstructed results as shown in figure 13(a). It can be observed that as the block number increases, the PSNR and SSIM of the reconstruction results show an increasing trend, and achieves the best performances when the block number is 8. Therefore, we chose 8 as the final number of blocks in the experiment. As the block number continues to

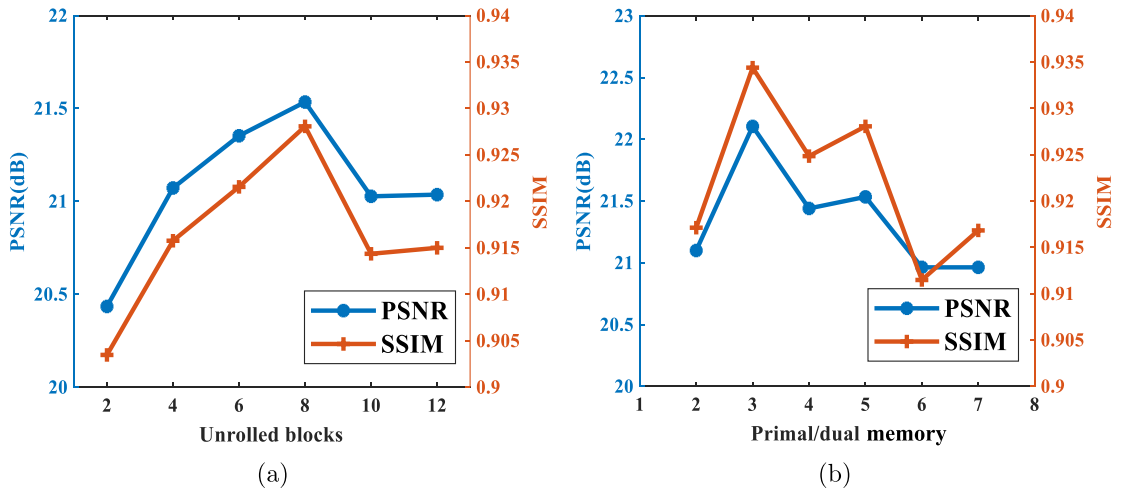

Figure 13. PSNR and SSIM yielded by different number of unrolled blocks and different number of extended primal/dual memory. (a) Average PSNR and SSIM calculated based test images with unrolled block number from 2 to 12. (b) Average PSNR and SSIM calculated based test images with extended primal/dual memory from 2 to 7.

increase, the effectiveness of the network does not continue to improve. We assume that the network may be too deep and it is difficult to obtain an acceptable convergence solution with the increased difficulties in training.

5.2. The number of extended primal and dual space

Following LPD, the primal space and the dual space, which corresponding to the reconstructed image space and the measured sinogram space are both extended to allow the STPDnet some 'memory' between the iterations. However, in different inverse problems, the optimal amount of expansion space should also be different, which has not been discussed in previous studies. We test the optimal amount of expansion primal/dual space for low count dynamic PET reconstruction problem when the block number is 8. The results are shown in figure 13(b). The network achieves the best performance when the number of extended primal/dual memory is 3.

5.3. Limitations

As one of the supervised learning methods, for STPDnet, a certain number of training data pairs is necessary. In order to obtain high-quality label data, a longer scan time or high level of dose is needed, which might introduce motion artifacts. However, unlike DeepPET, STPDnet does not necessitate extremely large data requirements due to the combination of physical projection process in the iterative framework. The domain shift between simulation data and clinical data is an inescapable challenge. Augmenting the quantity of simulated data, for instance, by varying the phantom shapes, count levels, and input function types, could theoretically alleviate the domain shift. Nevertheless, the resulting immense volume of data would considerably extend the training period, especially when using the deep unrolled method that includes a projection process within the network architecture. Consequently, training times may be unacceptably prolonged when handling large datasets.

The input of STPDnet is 3D sinogram data, namely 2D spatial dimension and 1D temporal dimension. However, dynamic PET reconstruction is a fully 4D problem and STPDnet cannot achieve fully 4D reconstruction due to the increased data size and memory requirement. Besides, TOF information is not considered in STPDnet. Although TOF information can greatly improve the quality of the reconstructed image, the space consumption brought by TOF information is not only unbearable for STPDnet, but also a problem for any deep learning algorithm based on sinogram data.

6. Conclusion

In this paper, we have developed a STPDnet for low count dynamic PET image reconstruction. The simulation results show that the STPDnet yields better performance in both quantitative analysis and bias variance analysis and show substantial noise reduction in both spatial domain and temporal domain compared with MLEM, KEM-ST, LPD and FBPnet. Real rat results show STPDnet better reconstruction performance in the low count situation, which makes the proposed method particularly suitable in whole-body dynamic imaging and parametric PET imaging that require extreme short frames and usually suffer from high level of noise. Future work will focus on kinetic parameters estimation of clinical patient data.

Acknowledgments

Animal experiments were approved by the experimental animal welfare and ethics review committee of Zhejiang University, and were performed in compliance with local legal requirements. This work is supported in part by the Talent Program of Zhejiang Province (No: 2021R51004), by the National Natural Science Foundation of China (No: U1809204, No: 62101488), by the Key Research and Development Program of Zhejiang Province (No: 2021C03029), and by NSF Grants: DMS2152961.

Data availability statement

The data cannot be made publicly available upon publication because they are owned by a third party and the terms of use prevent public distribution. The data that support the findings of this study are available upon reasonable request from the authors.

ORCID iDs

Rui Hu https://orcid.org/0000-0002-5322-2251

References

Adler J and Öktem O 2018 Learned primal-dual reconstruction *IEEE Trans. Med. Imaging* **37** 1322–32

Bauschke H H *et al* 2011 *Convex Analysis and Monotone Operator Theory in Hilbert Spaces* vol 408 (Berlin: Springer)

Boss D S, Olmos R V, Sinaasappel M, Beijnen J H and Schellens J H 2008 Application of PET/CT in the development of novel anticancer drugs *Oncologist* **13** 25–38

Brooks R A and Chiro G Di 1976 Statistical limitations in x-ray reconstructive tomography *Med. Phys.* **3** 237–40

Chambolle A and Pock T 2011 A first-order primal-dual algorithm for convex problems with applications to imaging *J. Math. Imaging Vis.* **40** 120–45

Chen S, Liu H, Hu Z, Zhang H, Shi P and Chen Y 2015 Simultaneous reconstruction and segmentation of dynamic PET via low-rank and sparse matrix decomposition *IEEE Trans. Biomed. Eng.* **62** 1784–95

Chen S, Liu H, Shi P and Chen Y 2015 Sparse representation and dictionary learning penalized image reconstruction for positron emission tomography *Phys. Med. Biol.* **60** 807

Cheng J-C *et al* 2021 Dynamic PET image reconstruction utilizing intrinsic data-driven HYPR4D denoising kernel *Med. Phys.* **48** 2230–44

Cui J, Qin Z, Chen S, Chen Y and Liu H 2019a Structure and tracer kinetics-driven dynamic PET reconstruction *IEEE Trans. Radiat. Plasma Med. Sci.* **4** 400–9

Cui J, Yu H, Chen S, Chen Y and Liu H 2019b Simultaneous estimation and segmentation from projection data in dynamic PET *Med. Phys.* **46** 1245–59

Cui J *et al* 2019c PET image denoising using unsupervised deep learning *Eur. J. Nucl. Med. Mol. Imaging* **46** 2780–9

Ehrhardt M J, Markiewicz P and Schönlieb C-B 2019 Faster PET reconstruction with non-smooth priors by randomization and preconditioning *Phys. Med. Biol.* **64** 225019

Feng D, Huang S-C and Wang X 1993 Models for computer simulation studies of input functions for tracer kinetic modeling with positron emission tomography *Int. J. Bio-Med. Comput.* **32** 95–110

Fessler J A 1994 Penalized weighted least-squares image reconstruction for positron emission tomography *IEEE Trans. Med. Imaging* **13** 290–300

Gong K, Berg E, Cherry S R and Qi J 2019 Machine learning in PET: from photon detection to quantitative image reconstruction *Proc. IEEE* **108** 51–68

Gong K *et al* 2019 EMnet: an unrolled deep neural network for PET image reconstruction *Med. Imaging 2019: Phys. Med. Imaging* **10948** 1203–8

Guazzo A and Colarieti-Tosti M 2021 Learned Primal dual reconstruction for PET *J. Imaging* **7** 248

Häggström I, Schmidtlein C R, Campanella G and Fuchs T J 2019 DeepPET: a deep encoder-decoder network for directly solving the PET image reconstruction inverse problem *Med. Image Anal.* **54** 253–62

Hashimoto F, Ohba H, Ote K, Kakimoto A, Tsukada H and Ouchi Y 2021 4D deep image prior: dynamic PET image denoising using an unsupervised four-dimensional branch convolutional neural network *Phys. Med. Biol.* **66** 015006

Hashimoto F, Ohba H, Ote K, Teramoto A and Tsukada H 2019 Dynamic PET image denoising using deep convolutional neural networks without prior training datasets *IEEE Access* **7** 96594–603

Hu R, Cui J, Yu C, Chen Y and Liu H 2023 STPDnet: spatial-temporal convolutional primal dual network for dynamic PET image reconstruction *IEEE 20th International Symposium on Biomedical Imaging (Cartagena, Colombia, April, 2023)* arXiv:2303.04667

Hu R and Liu H 2022 TransEM: residual swin-transformer based regularized PET image reconstruction *Medical Image Computing and Computer Assisted Intervention—MICCAI* vol 2022, pp 184–93

Hudson H M and Larkin R S 1994 Accelerated image reconstruction using ordered subsets of projection data *IEEE Trans. Med. Imaging* **13** 601–9

Kostakoglu L, Agress H Jr and Goldsmith S J 2003 Clinical role of FDG PET in evaluation of cancer patients *Radiographics* **23** 315–40

Li Y *et al* 2022 A deep neural network for parametric image reconstruction on a large axial field-of-view PET *Eur. J. Nucl. Med. Mol. Imaging* **50** 701–14

Lim H, Chun I Y, Dewaraja Y K and Fessler J A 2020 Improved low-count quantitative PET reconstruction with an iterative neural network *IEEE Trans. Med. Imaging* **39** 3512–22

Liu H, Gao F, Guo M, Xue L, Nie J and Shi P 2015 PET image reconstruction: mean, variance, and optimal minimax criterion *J. Phys. D: Appl. Phys.* **48** 155401

Machac J 2005 Cardiac positron emission tomography imaging *Semin. Nucl. Med.* **35** 17–36
Mehranian A and Reader A J 2020 Model-based deep learning PET image reconstruction using forward-backward splitting expectation-maximization *IEEE Trans. Radiat. Plasma Med. Sci.* **5** 54–64
Noh J, Fessler J A and Kinahan P E 2009 Statistical sinogram restoration in dual-energy CT for PET attenuation correction *IEEE Trans. Med. Imaging* **28** 1688–702
Onishi Y *et al* 2021 Anatomical-guided attention enhances unsupervised PET image denoising performance *Med. Image Anal.* **74** 102226
Qi J, Leahy R M, Cherry S R, Chatziioannou A and Farquhar T H 1998 High-resolution 3D Bayesian image reconstruction using the microPET small-animal scanner *Phys. Med. Biol.* **43** 1001
Reader A J, Corda G, Mehranian A, da Costa-Luis C, Ellis S and Schnabel J A 2020 Deep learning for PET image reconstruction *IEEE Trans. Radiat. Plasma Med. Sci.* **5** 1–25
Shepp L A and Vardi Y 1982 Maximum likelihood reconstruction for emission tomography *IEEE Trans. Med. Imaging* **1** 113–22
Wang B and Liu H 2020 FBP-Net for direct reconstruction of dynamic PET images *Phys. Med. Biol.* **65** 235008
Wang G 2018 High temporal-resolution dynamic PET image reconstruction using a new spatiotemporal kernel method *IEEE Trans. Med. Imaging* **38** 664–74
Wang G and Qi J 2012 Penalized likelihood PET image reconstruction using patch-based edge-preserving regularization *IEEE Trans. Med. Imaging* **31** 2194–204
Wang G and Qi J 2014 PET image reconstruction using kernel method *IEEE Trans. Med. Imaging* **34** 61–71
Wang H *et al* 2019 Accelerating MR imaging via deep Chambolle-Pock network *2019 41st Annual Int. Conf. of the IEEE Engineering in Medicine and Biology Society (EMBC)* pp 6818–21
Wernick M N, Infusino E J and Milosevic M 1999 Fast spatio-temporal image reconstruction for dynamic PET *IEEE Trans. Med. Imaging* **18** 185–95
Whiteley W, Panin V, Zhou C, Cabello J, Bharkhada D and Gregor J 2020 FastPET: near real-time reconstruction of PET histo-image data using a neural network *IEEE Trans. Radiat. Plasma Med. Sci.* **5** 65–77
Xie N *et al* 2020 Penalized-likelihood PET image reconstruction using 3D structural convolutional sparse coding *IEEE Trans. Biomed. Eng.* **69** 4–14
Yu F, Liu H, Hu Z and Shi P 2012 Graphics processing unit (GPU)-accelerated particle filter framework for positron emission tomography image reconstruction *J. Opt. Soc. Am.* A **29** 637–43
Zhang Z and Liu H 2019 Nonlocal total variation based dynamic PET image reconstruction with low-rank constraints *Phys. Scr.* **94** 065202
Zhou L, Schaefferkoetter J D, Tham I W, Huang G and Yan J 2020 Supervised learning with cyclegan for low-dose FDG PET image denoising *Med. Image Anal.* **65** 101770
Zhu B, Liu J Z, Cauley S F, Rosen B R and Rosen M S 2018 Image reconstruction by domain-transform manifold learning *Nature* **555** 487–92
Zubal I G, Harrell C R, Smith E O, Rattner Z, Gindi G and Hoffer P B 1994 Computerized three-dimensional segmented human anatomy *Med. Phys.* **21** 299–302