

LMPDNET: TOF-PET LIST-MODE IMAGE RECONSTRUCTION USING MODEL-BASED DEEP LEARNING METHOD

Chenxu Li Rui Hu Jianan Cui Huafeng Liu

State Key Lab of Modern Optical Instrumentation
Zhejiang University
Hangzhou, 310027, China

ABSTRACT

The integration of Time-of-Flight (TOF) information in the reconstruction process of Positron Emission Tomography (PET) yields improved image properties. However, implementing the cutting-edge model-based deep learning methods for TOF-PET reconstruction is challenging due to the substantial memory requirements. In this study, we present a novel model-based deep learning approach, LMPDNet, for TOF-PET reconstruction from list-mode data. We address the issue of real-time parallel computation of the projection matrix for list-mode data, and propose an iterative model-based module that utilizes a dedicated network model for list-mode data. Our experimental results indicate that the proposed LMPDNet outperforms traditional iteration-based TOF-PET list-mode reconstruction algorithms. Additionally, we compare the spatial and temporal consumption of list-mode data and sinogram data in model-based deep learning methods, demonstrating the superiority of list-mode data in model-based TOF-PET reconstruction.

Index Terms— TOF-PET reconstruction, list-mode, model-based deep learning

1. INTRODUCTION

Incorporating Time-of-Flight (TOF) information in Positron Emission Tomography (PET) imaging enhances image quality and bolsters lesion detectability[7, 14]. The main methods for TOF-PET reconstruction include Ordered Subset Expectation Maximization (OSEM)[13, 10], Maximum A Posteriori (MAP)[2], Primal-Dual Hybrid Gradient (PDHG)[12], etc. These algorithms are suitable for both sinogram data and list-mode data. However, these iterative-based algorithms suffer from computational inefficiency and may produce reconstructed images with very poor quality when applied to the data of low counts.

In recent times, machine learning has become a widely-utilized technique used in the field of medical image reconstruction due to its high computational efficiency and data-driven characteristic. Numerous studies have demonstrated the efficacy of model-based deep learning methods

in reconstructing high-quality PET images[9, 11, 8], and the model-based approach also provides interpretability to deep learning[5]. However, there are few model-based deep learning methods for TOF-PET reconstruction. On one hand, the storage requirements of the system matrix corresponding to the sinogram combined with TOF information will be even greater. On the other hand, starting from list-mode data, how to integrate the system matrix of list-mode data into the iterative framework is a problem worth discussing. For a 2D image with a dimension of 128×128 and sinogram with a dimension of $357 \times 224 \times 17$ (rad \times view \times tofbin), the system matrix occupies ~ 83 GB of video memory. What's worse, as the increase of axial field of view, its memory usage will also increase dramatically. This severely limits the application of model-based learning methods on TOF-PET reconstruction. The reconstruction algorithm based on list-mode data can discard the fixed format like sinogram. It ensures that the data are all valid information, rather than a large number of "empty bins" appearing in the sinogram in TOF-PET, avoiding the waste of memory. The memory occupied by the list-mode data depends on the number of coincident events detected in one scan, regardless of whether there is TOF information or not. For the example mentioned above, if the count of coincident events is $1e5$, the projection matrix (corresponding to the system matrix) occupies only ~ 6 GB of video memory.

In this work, we proposed the LMPDNet, i.e. List-mode Primal Dual Net, a novel model-based deep learning approach for list-mode data to improve the TOF-PET reconstruction. We first address the issue of real-time calculation of the projection matrix during iterative training. Based on the projection matrix calculation method proposed by Joseph[6], we used CUDA to compute the projection of each response line (LOR) in parallel to achieve acceleration. Then, based on the Learned Primal-Dual method[1, 4], the LMPDNet was proposed. Simulation experiments showed that our method was superior to the current mainstream TOF-PET list-mode reconstruction method[10, 12]. To the best of our knowledge, this is the first unrolled model-based deep learning method for TOF-PET list-mode reconstruction.

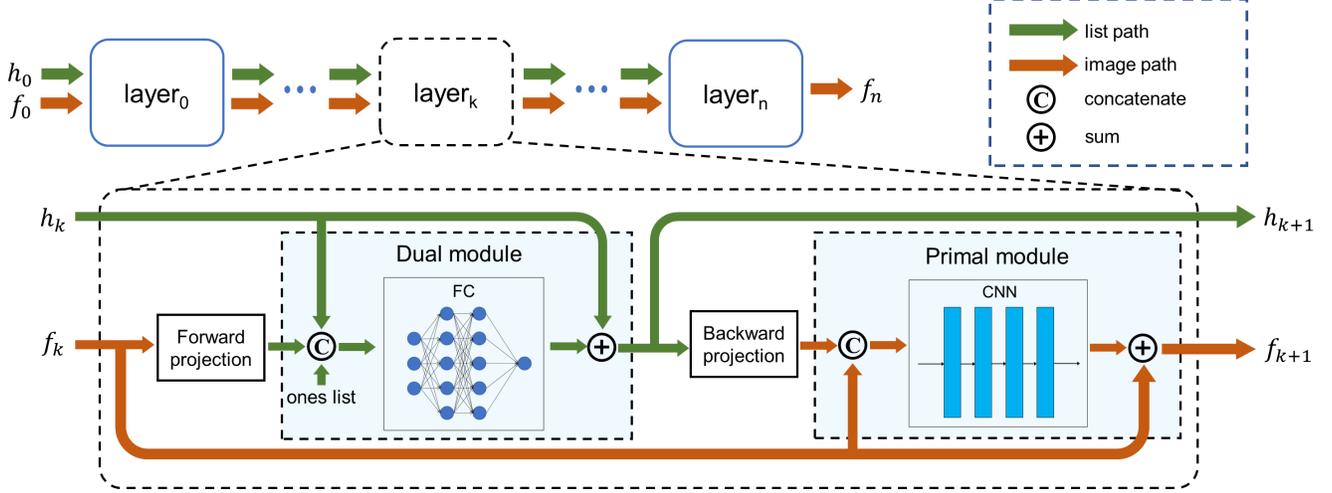


Fig. 1. Overall framework of LMPDNet

2. METHOD

2.1. Learned Primal-Dual

Mathematically, the TOF-PET reconstruction problem from list-mode can be formulated as:

$$g = Pf + noise \quad (1)$$

where $f \in X$ (image domain) and $g \in Y$ (measurement domain), and P is the forward projection matrix that mapping $X \rightarrow Y$ with TOF information.

A general approach to solve (1) is to minimize the sum of the negative log-likelihood function $\mathcal{L}(Pf, g)$ and penalized regularization term $\mathcal{S}(f)$:

$$\min_{f \in X} [\mathcal{L}(Pf, g) + \lambda \mathcal{S}(f)] \quad (2)$$

where the λ is the regularization parameter. This is a large-scale optimization problem. A general framework called the Learned Primal-Dual [1] combining deep learning with PDHG was proposed for solving optimization problems of the form similar to (2). PDHG is based on the use of forward and backward projections, and is capable of solving problems with complex constraints, such as positivity and sparsity. Shown in Algorithm 1, The main idea of Learned Primal-Dual is to unroll the PDHG algorithm, keeping the forward projection operator \mathcal{T} and its adjoint projection operator \mathcal{T}^* , and replacing the proximal operator in the PDHG algorithm with the parameterized operator Γ_{θ^d} and Λ_{θ^p} , and the parameters θ^d and θ^p are learned from the training data. h_k and f_k represents the data in measurement domain and image domain respectively. For list-mode data, the operator \mathcal{T} is the projection matrix P .

2.2. Overall Framework

The overall framework of the LMPDNet is shown in Figure 1. Like the Learned Primal-Dual reconstruction method, the

Algorithm 1 Learned Primal-Dual

- 1: Initialize $f_0 (= 0), h_0 (= 0)$
- 2: **for** $k = 1, \dots, K$ **do**
- 3: $h_k \leftarrow \Gamma_{\theta^d}(h_{k-1}, \mathcal{T}(f_{k-1}), g)$
- 4: $f_k \leftarrow \Lambda_{\theta^p}(f_{k-1}, \mathcal{T}^*(h_k))$
- 5: **return** f_K
- 6: **end for**

structure of the proposed method is formed by concatenating multiple identical layers. The inputs are list-mode space input $h_0 \in \mathbb{R}^n$ and image space input $f_0 \in \mathbb{R}^{W \times H}$ initialized to zero, where n is the length of list-mode data, W and H are width and height of the image. Each layer consists of four parts: Forward projection, Dual module, Backward projection and Primal module.

Forward/backward projection: For sinogram data, the operator \mathcal{T} and \mathcal{T}^* in Algorithm 1 is the system matrix G and its adjoint matrix G^* , which can be precomputed. Unlike sinogram data, the corresponding projection matrix P of list-mode data is different for each set of detection data and it cannot be precomputed. This is because the sequence of coincidence events in list-mode data is random and disordered, and its length is not fixed. Therefore, for each set of list-mode data, its projection matrix needs to be calculated in real-time before being input to the neural network.

The projection result of a LOR line is the integral of the activity on this line:

$$h(i) = \int_{LOR_i} \epsilon f(x, y) ds \quad (3)$$

where the ϵ is the TOF weight, $i \in [1, n]$ represents the i -th LOR, $f(\cdot)$ is the activity distribution and $x \in [1, W], y \in [1, H]$ represents image coordinates. ds represents the line differentiation of the LOR. The contribution of each pixel to

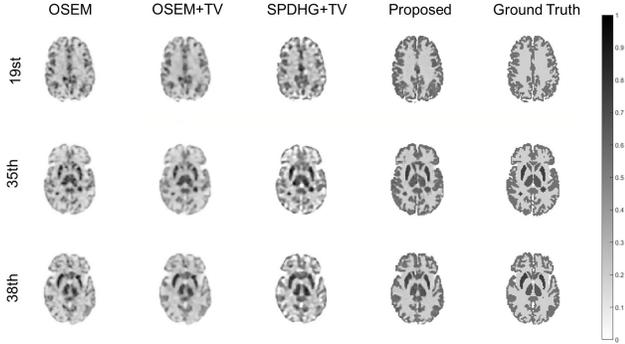


Fig. 2. Reconstruction result of OSEM, SPDHG+TV and proposed method. The 19st, 35th, and 38th slices in phantom are selected for display.

the LOR varies linearly with the position, so the result of the line integration can be expressed as the product of the length of the line segment and the midpoint activity value of the line segment:

$$h(i) = \sum_{y=1}^H \epsilon f(s_{mid}) \cdot \Delta s \quad (4)$$

where Δs is the length of line segment. The midpoint activity value can be calculated by linear interpolation between two adjacent pixels, then:

$$f(s_{mid}) = \rho f(x_l, y) + (1 - \rho) f(x_r, y) \quad (5)$$

ρ is the coefficient of linear interpolation, x_l and x_r are the coordinates to the left and right of the midpoint. Recall that a single element P_{ij} in the projection matrix $P \in [n, W \times H]$ represents the degree of influence caused by the i -th LOR on the j -th pixel. The result of multiplying P_i (the i -th row in P) with $f(\cdot)$ (the image resized as a column vector) is $h(i)$. That is:

$$P_i \cdot f(\cdot) = \sum_{y=1}^H \epsilon [\rho f(x_l, y) + (1 - \rho) f(x_r, y)] \Delta s \quad (6)$$

Dividing the terms relating to f , then the calculation result of the element value in the projection matrix is:

$$P_{ij} = \epsilon \cdot \rho \cdot \Delta s \quad (7)$$

There is no data communication between the calculations of different LOR projections, and the granularity of each LOR projection calculation is relatively fine. Therefore, the projection matrix calculation can be accelerated by CUDA parallel computing. It is worth noting that the projection matrix can be used directly for the subsequent backpropagation of the neural network after it has been computed by requesting memory in the

Table 1. PSNR and SSIM (mean \pm std) of various methods and counts

Methods	Counts=3e5		Counts=1e5	
	PSNR	SSIM	PSNR	SSIM
OSEM	17.68 \pm 1.26	0.84 \pm 0.03	16.96 \pm 1.22	0.82 \pm 0.04
OSEM+TV	18.26 \pm 1.25	0.85 \pm 0.03	17.81 \pm 1.24	0.84 \pm 0.03
SPDHG+TV	18.37 \pm 1.17	0.85 \pm 0.03	15.44 \pm 1.10	0.80 \pm 0.04
Proposed	23.47\pm0.49	0.94\pm0.01	21.52\pm0.44	0.92\pm0.01

GPU, without the need to copy the data from CPU memory to GPU memory.

Dual Module: The input of the Dual Module is h_k and the forward projection result of f_k . Concatenate two input lists and an all-ones list, and get $\tilde{h} \in \mathbb{R}^{n*3}$. The all-ones list is to serve as the reference, i.e. g in Learned Primal-Dual, for the forward projection result, which is like using the measured original sinogram as the reference during the sinogram projection process. Considering that there is no correlation among the elements of a column in \tilde{h} , we need to pay more attention to the connection between the three values of each row. A fully connected (FC) neural network with three input features and one output feature was used, and the dimension of length n becomes the “channel” dimension. Finally, the output result is used as the residual and the list obtained from the previous iteration h_k is summed to obtain the output result of Dual Module.

Primal Module: Primal Modules are similar in structure to Dual modules. The difference is that the input and output of this module are not lists, but images. The input of the Primal Module is the concatenation of f_k and the backward projection result of h_{k+1} . The Primal module uses Convolutional Neural Networks (CNNs) with PReLU as activation function.

3. EXPERIMENTS

3.1. Datasets

The Zubal phantom[15] with size of $128 \times 128 \times 40$ was used as a starting point for the simulated data. We inserted several hot spheres of radius 2mm to 4mm at random positions in phantom to simulate the tumors and obtained 12 different phantoms. Then we got 480 different 2D phantoms of size

Table 2. Comparison of video memory occupation and projection time consuming

	sinogram	list-mode
Memory Occupation	\sim 87GB	\sim 20GB
Forward Projection	364ms	256ms
Backward Projection	405ms	246ms
Compute Projection Matrix	-	157ms

128×128. A three compartment model was used along with the Feng’s input function to simulate the activity map[3]. In order to simulate the list-mode data, we first obtained the system matrix for the specific PET scanner using (6) . The system matrix was multiplied by the activity map to obtain the original sinograms. Next, the sinograms were scaled down according to the count rate, the total count was set to 3e5 and 1e5 respectively. Then 15% Poisson random noise was added to sinograms. Finally, sinograms were converted to list-mode data by rounding down the value of each bin in the sinograms to the number of events in its corresponding position, and using each event as a row in the list-mode data. At this point, we had 480 pairs of activity map and list-mode data. Among them, 400 pairs of data were selected as the training set, 40 pairs as the validation set, and 40 pairs as the test set.

The system we used was a simulated classic cylindrical PET scanner, which included 28 modules with 16 crystals each, having a width of 4 mm in the radial direction. Set 17 TOF-bins on each LOR, each TOF-bin is 15mm long and the time resolution of TOF was set as 400 ps. Thus the dimension of the sinogram was 357×224×17. It is worth noting that axial information in the detection system was not required since the experiment focused on 2D reconstruction of PET images.

3.2. Training Details

The number of network iteration layers was set to 7. Each Dual-Net consisted of two hidden layers with 32 features, and Primal-Net consisted of 4 convolutional layers, the number of channels in the first three layers is set to 64, except for the last layer where the number of output channels is 1. The MSE Loss between the output images f_n and the ground truth was chosen as the loss function, and the learning rate was set to 1e-5. The training epochs are 500.

3.3. Results

To quantitatively evaluate the quality of the reconstruction results, we calculated the Peak Signal-to-Noise Ratio (PSNR) and Structural Similarity Index Measure (SSIM) between the reconstructed images and the ground truth. The comparison methods included the traditional OSEM algorithm, OSEM with TV regularization, and the SPDHG algorithm with TV regularization. The traditional algorithms all had 30 iterations and were divided into 4 subsets. The results are presented in Figure 2 and Table 1. The findings suggest that, for both 3e5 counts and 1e5 counts, the proposed method outperforms the traditional model-based reconstruction algorithm in terms of both quantitative and qualitative analysis. Notably, as no other model-based deep learning method of list-mode reconstruction currently exists, we did not compare our proposed method with other learning-based methods.

In Table 2, we conducted a comparison between using sinogram and list-mode data for reconstruction under the ex-

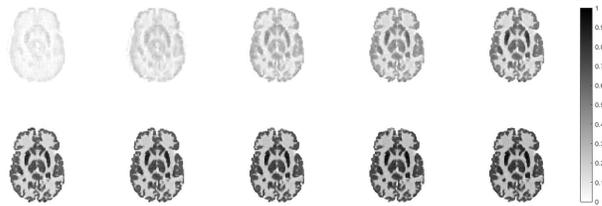


Fig. 3. The output of each layer of the proposed method. The first row from left to right represents layers 1 to 5 and the second row from left to right represents layers 6 to 10.

perimental settings described above. Specifically, we compared the video memory occupation and the time required for each forward/backward projection. The results indicated that using list-mode data as the starting point resulted in lower memory consumption and less projection time compared to using sinogram data. Due to the need of calculating the projection matrix in real-time, the method based on list-mode data consumed a part of the time when each new set of data was fed into the neural network. However, this time consumption only occurs before the new set of data enters the overall reconstruction process, not before each forward/backward projection. And after the acceleration of parallel computing, this time-consuming was small and negligible.

3.4. Ablation Study

To balance the effectiveness of the proposed method against its memory usage, we need to investigate the optimal number of iterative layers to be employed. Initially, we set the number of layers to 10, and the output of each layer was examined and compared with the ground truth. The results are presented in Figure 3. Our analysis revealed that the output of each layer became increasingly more accurate as the number of iterative layers increased, eventually stabilizing at around the 7th layer. Based on these findings, we set the number of layers to 7 for our experiments.

4. CONCLUSION

This paper proposed a model-based deep learning method for TOF-PET reconstruction from list-mode data called LMPDNet, which was the first application of model-based deep learning to TOF-PET reconstruction. The results showed that the proposed method outperformed traditional iteration-based algorithms in terms of reconstruction quality and outperformed reconstruction from sinogram in terms of spatial and temporal consumption.

5. REFERENCES

- [1] Jonas Adler and Ozan Öktem. Learned primal-dual reconstruction. *IEEE Transactions on Medical Imaging*, 37(6):1322–1332, 2018.
- [2] Bing Bai, Yanguang Lin, Wentao Zhu, Ran Ren, Quanzheng Li, Magnus Dahlbom, Frank DiFilippo, and Richard M Leahy. Map reconstruction for fourier rebinned tof-pet data. *Physics in Medicine & Biology*, 59(4):925, 2014.
- [3] Dagan Feng, Sung-Cheng Huang, and Xinmin Wang. Models for computer simulation studies of input functions for tracer kinetic modeling with positron emission tomography. *International Journal of Bio-Medical Computing*, 32(2):95–110, 1993.
- [4] Alessandro Guazzo and Massimiliano Colarieti-Tosti. Learned primal dual reconstruction for pet. *Journal of Imaging*, 7(12):248, 2021.
- [5] Chang Min Hyun, Seong Hyeon Baek, Mingyu Lee, Sung Min Lee, and Jin Keun Seo. Deep learning-based solvability of underdetermined inverse problems in medical imaging. *Medical Image Analysis*, 69:101967, 2021.
- [6] Peter M Joseph. An improved algorithm for reprojecting rays through pixel images. *IEEE Transactions on Medical Imaging*, 1(3):192–196, 1982.
- [7] Joel S Karp, Suleman Surti, Margaret E Daube-Witherspoon, and Gerd Muehlehner. Benefit of time-of-flight in pet: experimental and clinical results. *Journal of Nuclear Medicine*, 49(3):462–470, 2008.
- [8] Kyungsang Kim, Dufan Wu, Kuang Gong, Joyita Dutta, Jong Hoon Kim, Young Don Son, Hang Keun Kim, Georges El Fakhri, and Quanzheng Li. Penalized pet reconstruction using deep learning prior and local linear fitting. *IEEE Transactions on Medical Imaging*, 37(6):1478–1487, 2018.
- [9] Abolfazl Mehranian and Andrew J Reader. Model-based deep learning pet image reconstruction using forward-backward splitting expectation-maximization. *IEEE Transactions on Radiation and Plasma Medical Sciences*, 5(1):54–64, 2020.
- [10] Guillem Pratx, Suleman Surti, and Craig Levin. Fast list-mode reconstruction for time-of-flight pet using graphics hardware. *IEEE Transactions on Nuclear Science*, 58(1):105–109, 2010.
- [11] Andrew J Reader, Guillaume Corda, Abolfazl Mehranian, Casper da Costa-Luis, Sam Ellis, and Julia A Schnabel. Deep learning for pet image reconstruction. *IEEE Transactions on Radiation and Plasma Medical Sciences*, 5(1):1–25, 2020.
- [12] Georg Schramm and Martin Holler. Fast and memory-efficient reconstruction of sparse poisson data in list-mode with non-smooth priors with application to time-of-flight pet. *Physics in Medicine & Biology*, 2022.
- [13] Donald L Snyder and David G Politte. Image reconstruction from list-mode data in an emission tomography system having time-of-flight measurements. *IEEE Transactions on Nuclear Science*, 30(3):1843–1849, 1983.
- [14] Suleman Surti and Joel S Karp. Update on latest advances in time-of-flight pet. *Physica Medica*, 80:251–258, 2020.
- [15] I George Zubal, Charles R Harrell, Eileen O Smith, Zachary Rattner, Gene Gindi, and Paul B Hoffer. Computerized three-dimensional segmented human anatomy. *Medical Physics*, 21(2):299–302, 1994.