UC Davis

Title

Deep Kernel Representation for Image Reconstruction in PET

Permalink

Journal

ISSN

Authors

Li, Siqi
Wang, Guobao

Publication Date

DOI

Peer reviewed

Deep Kernel Representation for Image Reconstruction in PET

Siqi Li,

Department of Radiology, University of California Davis Health, Sacramento, CA 95817, USA

Guobao Wang

Department of Radiology, University of California Davis Health, Sacramento, CA 95817, USA

Abstract

Image reconstruction for positron emission tomography (PET) is challenging because of the ill-conditioned tomographic problem and low counting statistics. Kernel methods address this challenge by using kernel representation to incorporate image prior information in the forward model of iterative PET image reconstruction. Existing kernel methods construct the kernels commonly using an empirical process, which may lead to unsatisfactory performance. In this paper, we describe the equivalence between the kernel representation and a trainable neural network model. A deep kernel method is then proposed by exploiting a deep neural network to enable automated learning of an improved kernel model and is directly applicable to single subjects in dynamic PET. The training process utilizes available image prior data to form a set of robust kernels in an optimized way rather than empirically. The results from computer simulations and a real patient dataset demonstrate that the proposed deep kernel method can outperform the existing kernel method and neural network method for dynamic PET image reconstruction.

Keywords

Dynamic PET; deep kernel learning; image reconstruction; kernel methods; neural networks

I. Introduction

POSITRON emission tomography (PET) is an imaging modality for quantitatively measuring biochemical and physiological processes *in vivo* by using a radiotracer [1]. Image reconstruction for PET is challenging due to the ill-conditioned tomographic problem and low-counting statistics (high noise) of PET data [2], for example, in dynamic PET imaging where short time frames are used to monitor rapid change in tracer distribution.

Among different methods of PET image reconstruction, the kernel methods (e.g., [3]–[9]) address the noise challenge by uniquely integrating image prior information into the forward model of PET image reconstruction through a kernel representation framework [3]. Image prior information may come from composite time frames of a dynamic PET scan [3], or from anatomical images (e.g., magnetic resonance (MR) images [4], [5] in integrated PET/MRI). The kernel methods can be easily implemented with the existing

sqlli@ucdavis.edu .

expectation-maximization (EM) algorithm and have demonstrated substantial image quality improvement as compared to other methods [3]–[5].

In the existing kernel methods, a kernel representation is commonly built using an empirical process for defining feature vectors and manually selecting method-associated parameters [3]. However, such an experience-based parameter tuning and optimization approach often leads to suboptimal performance. In this paper, we first describe the equivalence between the kernel representation and a trainable neural network model. Based on this connection, we then propose a deep kernel method that learns the trainable components of the neural network model from available image data to enable a data-driven automated learning of an improved kernel model. The learned kernel model is then applied to tomographic image reconstruction and is expected to outperform existing kernel models that are empirically defined.

There are many ongoing efforts in the field to explore deep learning with neural networks for PET image reconstruction, see recent review articles, e.g., [10]–[15]. Deep neural networks have been proposed for direct mapping from the projection domain to the image domain (e.g., [16]) but the models are so far mainly practical for 2D data training. By unrolling an iterative tomographic reconstruction algorithm, model-based deep-learning reconstruction (e.g., [17], [18]) represents a promising direction. One limitation of this method is that it requires pre-training using a large number of data sets and involves projection data in the iterative training process, which is computationally intensive. Alternatively, neural networks can be used as "deep image prior" for image representation in iterative reconstruction, e.g. by the convolutional neural network (CNN) model [19]–[24]. It has the advantage of being directly applicable to single subjects. The resulting reconstruction problem, however, is nonlinear and is often complex and challenging to optimize.

Different from these methods that utilize pure neural networks, the proposed deep kernel method combines deep neural networks into the kernel framework [3] to form a novel way for tomographic image representation. The method has a unique advantage that once the model is trained with neural networks, the unknown kernel coefficient image remains linear in the model and is therefore easy to be reconstructed from PET data. It does not necessarily require a large data set for training but is directly applicable to single-subject learning and reconstruction, e.g., in dynamic PET, as will be demonstrated in this paper.

The rest of this paper is organized as follows. Section II introduces the background materials of the kernel method for PET image reconstruction. Section III describes the generalized theory of the proposed deep kernel method that derives a data-driven automated learning of an improved kernel method. We then present a computer simulation study in Section IV and a real patient data study in Section V to demonstrate the improvement of the proposed method over existing methods. Finally discussions and conclusions are drawn in Section VI and VII.

II. Background

A. PET Image Reconstruction

PET projection data $y = \{y_i\}$ can be well modeled as independent Poisson random variables using the log-likelihood function [2],

$$L(y \mid x) = \sum_{i=1}^{N} y_i \log \bar{y}_i - \bar{y}_i - \log y_i!, \tag{1}$$

where i denotes the detector index and N is the total number of detector pairs. The expectation of the projection data, $\bar{y}$, is related to the unknown image x through

$$\bar{y} = Px + r, \tag{2}$$

where P is the detection probability matrix for PET and includes normalization factors for scanner sensitivity, scan duration, deadtime correction and attenuation correction. r is the expectation of random and scattered events [2].

The maximum likelihood (ML) estimate of the image x is found by maximizing the Poisson log-likelihood,

$$\hat{x} = \arg \max_{x \geq 0} L(y \mid x). \tag{3}$$

A common way of seeking the solution of (3) is to use the EM algorithm [25].

B. Kernel Methods for PET Reconstruction

The kernel methods describe the image intensity x_j at the pixel j as a linear representation of kernels [3],

$$x_j = \sum_{l \in N_j} \alpha_l \kappa(f_j, f_l), \quad j, l = 1, \cdots, n_p \tag{4}$$

where N_j defines the neighborhood of pixel j and n_p is the total number of image pixels. f_j and f_l are the feature vectors extracted from image priors for pixel j and pixel l, respectively. α_l is the kernel coefficient at pixel l. $\kappa(\cdot, \cdot)$ is the kernel function that defines a weight between pixel j and pixel l. A popular choice of $\kappa(\cdot, \cdot)$ is the radial Gaussian kernel,

$$\kappa(f_j, f_l) = \exp\left(-\frac{\|f_j - f_l\|^2}{2\sigma^2}\right). \tag{5}$$

with σ being the kernel parameter. The equivalent matrix-vector form of (4) is

$$x = K\alpha \tag{6}$$

with the (j, l)th element of the square kernel matrix K being $\kappa(f_j, f_l)$.

The kernel coefficient image α is then estimated from the projection data y by maximizing the log-likelihood L,

$$\hat{\alpha} = \arg \max_{\alpha \geq 0} L(y \mid K\alpha),$$

(7)

which can be solved using the kernelized EM algorithm [3],

$$\alpha^{n+1} = \frac{\alpha^n}{K^T P^T 1_N} \cdot \left(K^T P^T \frac{y}{PK\alpha^n + r} \right),$$

(8)

where n denotes the iteration number and the superscript "T" denotes matrix transpose. 1_N is a vector with all elements being 1. Once α is estimated, the final PET activity image x is given by $\hat{x} = K\hat{\alpha}$.

Note that in practice, a normalized kernel matrix

$$\bar{K} = \text{diag}^{-1}[K1_N]K$$

(9)

is commonly used for better performance [3]. The (j,l)th element of $\bar{K}$ is equivalent to

$$\kappa(f_j, f_l) = \frac{\exp\left(-\frac{\|f_j - f_l\|^2}{2\sigma^2}\right)}{\sum_{l' \in N_j} \exp\left(-\frac{\|f_j - f_{l'}\|^2}{2\sigma^2}\right)}, \quad l \in N_j.$$

(10)

The neighborhood N_j of pixel j can be defined by its k-nearest neighbors (kNN) [29] to make K sparse. The feature vector f_j is usually set to the intensity values of the image prior at pixel j and the kernel parameter σ is chosen empirically, e.g. $\sigma = 1.0$.

III.   Proposed Deep Kernel Method

A.   Kernel Representation as Neural Networks

We first describe the kernel representation using a neural network description illustrated in Fig. 1. The construction of kernel representation is decomposed into two main modules: (1) feature extraction and (2) pairwise attention.

Denote the image prior data by Z which consists of n_z prior images, each with n_p pixels. The feature extraction module is to extract a feature vector f of length n_f for each image pixel from Z,

$$f_j = \Psi_j(Z),$$

(11)

where Ψ denotes the feature extraction operator, for example, a convolutional neural network. The extraction of conventional intensity-based features is equivalent to a $1 \times 1 \times 1$ convolution operations on Z (if the images are 3D). This step provides a feature data F of size $n_p \times n_f$.

The pairwise attention module first calculates the similarity between pixel j and its neighboring pixels that are specified by a pre-determined neighborhood N_j,

$$s_{jl} = -\frac{\|f_j - f_l\|^2}{2\sigma^2}, l \in N_j. \tag{12}$$

Note that here only k neighbors are selected for each pixel j, e.g., using the Euclidean distance-based kNN algorithm [3], [29]. This leads to a similarity data S of size $n_p \times k$. Pairwise weights are then calculated from S using

$$w_{jl} = \frac{\exp(s_{jl})}{\sum_{l' \in N_j} \exp(s_{jl'})}, \tag{13}$$

generating a weight data W of size $n_p \times k$. Here $\sum_{l \in N_j} w_{jl} = 1$. This type of weight calculation is also called *softmax* in neural networks and can be directly explained as a pairwise attention mechanism [26], [27]. w_{jl} is the attention weight of other "key" pixels $\{l\}$ as compared to the "query" pixel j.

The final step reshapes W using the neighborhood indices defined by $\{N_j\}_{j=1}^{n_p}$ to generate a sparse matrix, which is equal to the normalized kernel matrix defined in (9). Each row of the kernel matrix is of the size $n_p \times 1$ and can be displayed as an image, which can also be understood as an attention map for the corresponding pixel in α.

B. Deep Kernel Model

Integrating all the neural network components in Fig. 1 together, we have the following deep kernel model to represent a PET image x,

$$x = \mathcal{K}(\theta; Z)\alpha, \tag{14}$$

where $\mathcal{K}(\theta; Z)$ denotes the equivalent neural network model of K with the image prior data Z as the input and θ collecting any model parameters that are trainable.

The deep kernel model is nonlinear with respect to θ and Z but remains linear with respect to the kernel coefficient image α. While this model shares the spirit of using attention with the nonlocal neural network [27], the linearity of α makes it unique and more suitable for tomographic image reconstruction problems. Once θ is determined, α can be easily reconstructed from the projection data y using the kernelized EM algorithm in (8).

In conventional kernel methods, θ is equivalent to be determined empirically, which does not explore the full potential of the kernel method. For example, intensity-based features are commonly used for f. However, convolutional neural network-derived features can be more informative [28]. In this paper, we exploit the capability of deep learning to train an optimized feature set for generating K from available image prior data based on the proposed deep kernel model.

C. Deep Kernel Learning

The deep kernel learning problem is formulated using the observation that in the kernelized image model (6), x is usually a clean version of a if a is noisy. This inspires the following use of the denoising autoencoder framework [30] to train the model parameters of $\mathcal{K}(\theta; Z)$,

$$\hat{\theta} = \arg\min_{\theta} \sum_{q=1}^{n_{\text{tr}}} \left\| I_q - \mathcal{K}(\theta; Z)\tilde{I}_q \right\|^2, \tag{15}$$

where I_q denotes the qth high-quality image in the training dataset and $\tilde{I}_q$ is a corrupted version of I_q. n_{tr} is the total number of training image pairs. In PET, I_q and $\tilde{I}_q$ can be obtained from high count data and low-count data, respectively.

The deep kernel model can be pretrained using a large number of patient scans (large n_{tr}) if such a training dataset is available. It can also be trained online for single subjects (small n_{tr}) without pretraining, as described below.

D. Single-Subject Deep Kernel Method for Dynamic PET

In dynamic PET, the image prior data Z may consist of several composite images $\{z_m\}_{m=1}^{n_Z}$ where n_Z is the number of composite frames. These images are reconstructed from the rebinned long-scan projection data $\{y_m^{\text{reb}}\}_{m=1}^{n_Z}$ and may have good image quality due to the relatively high count level of a composite frame. For example, a one-hour dynamic ^{18}F-fluorodeoxyglucose (FDG) PET scan can be divided into three composite frames, each of 20 minutes [3]. The composite image prior has been used in the standard kernel methods for constructing the kernel matrix empirically. Here we use it to train an improved kernel model adaptive to a single subject.

The single-subject deep kernel learning problem for dynamic PET is constructed using the following optimization criterion,

$$\hat{\theta} = \arg\min_{\theta} \sum_{m=1}^{n_Z} \left\| z_m - \mathcal{K}(\theta; Z)\tilde{z}_m \right\|^2, \tag{16}$$

where the corrupted image $\tilde{z}_m$ can be obtained from the reconstruction of the low-count projection data which are downsampled from y_m^{reb} using a count reduction factor d (e.g. $d = 10$). Once θ is trained, the learned kernel model is then used to reconstruct all the dynamic frames of the scan frame-by-frame using the kernel EM algorithm in (8).

In theory, both the feature extraction and pairwise attention modules in the neural network model (Fig. 1) are trainable. As a proof of concept, in this work we only train the feature extraction operator Ψ, while the pairwise attention module is calculated using (12) and (13) as used in the conventional kernel method [3].

E.   Model Structure of Feature Extraction

The proposed method is applicable to different neural network architectures if they are suitable for image representation. Here a popular residual U-net architecture [19], as illustrated in Fig. 2, is used for the feature extraction module Ψ. The network is available in both 2D and 3D versions for learning 2D and 3D images, respectively. It consists of the following operations: 1) 3×3 (×3) 2D (3D) convolutional layer, 2) 2D (3D) batch normalization (BN) layer, 3) leaky rectified linear unit (LReLU) layer, 4) 3×3 (×3) convolutional layer with stride 2×2 (×2) for down-sampling, 5) 2×2 (×2) bilinear (trilinear) interpolation layer for up-sampling, 6) identity mapping layer that adds feature maps from left-side encoder path to the right-side decoder path. In addition, a ReLU layer is used before the output in order to satisfy the non-negative constraint on the last feature map.

IV.   COMPUTER SIMULATION VALIDATION

A.   Simulation Setup

Dynamic ^{18}F-FDG PET scans were simulated for a GE DST whole-body PET scanner in two-dimensional mode using a Zubal head phantom shown in Fig. 3a. The phantom is composed of gray matter, white matter, blood pools and a tumor (15 mm in diameter). A early 20-minute dynamic scan was divided into 63 time frames: 30×2s, 12 ×5s, 6 ×30s, and 15×60s. The pixel size is 3×3 mm^2 and the image size is 111×111.

The time activity curves of different regions are shown in Fig. 3(b–c). An attenuation map was simulated with a constant linear attenuation coefficient assigned in the whole brain. Dynamic images were first forward projected to generate noise-free sinograms. Poisson noise was then introduced. A 20% uniform background was included to account for mean random and scatter events. The expected total number of events over 20 min was 20 million. Twenty noisy realizations were simulated and each was reconstructed independently for comparison.

B.   Reconstruction Methods

The simuated dynamic data were reconstructed using four different methods: (1) standard ML-EM reconstruction; (2) existing kernel EM [3]; (3) the deep image prior (DIP) reconstruction method [19] as a recent representative of nonlinear neural network-based reconstruction methods; and (4) proposed deep kernel method with single-subject online training of the feature extraction module. The deep kernel method was trained separately for each of the 20 noisy data realizations. All reconstructions were run for 200 iterations with a uniform initial image.

The image priors for the kernel methods were the composite images obtained from four composite frames, each with 5 min scan. For the conventional kernel method, pixel intensity values extracted from the composite images $\{z_m\}$ were used to form the feature vector f for generating the kernel matrix K using kNN with $k = 48$ in the same way as used in [3].

The DIP method was implemented using the alternating direction method of multipliers (ADMM) in a way similar to [19] but was adapted to use the composite image prior data

as the input of the U-net. Within each outer iteration, 4 iterations were used for solving the penalized-likelihood image reconstruction problem and 50 iterations were used for the image-domain DIP learning. These settings were empirically optimized for obtaining stable results across different time frames according to the image mean squared error (MSE) in our experiments. The effect of the ADMM hyper-parameter ρ was also investigated and $\rho = 5 \times 10^{-6}$ was chosen to obtain nearly optimal image MSE.

In the deep kernel method, the low-count images $\{\tilde{z}_m\}$ were obtained by using one-tenth of the counts in each composite frame z_m. ML-EM was used to reconstruct the image pair $\{\tilde{z}_m\}$ and $\{z_m\}$ for training. The k in kNN for defining the neighborhood $\{N_j\}$ was set to be 200 for optimized image MSE performance. For implementation, the tomographic reconstruction step was implemented in MATLAB and the deep kernel training step was implemented in PyTorch, both on a PC with an Intel i9-9920X CPU with 64GB RAM and a NVIDIA GeForce RTX 2080Ti GPU. Three hundred iterations were used for the training step with the learning rate set to 10^{-3}. The Kaiming initialization method [31] was used for each convolutional layer and uniform initialization was used for each BN layer.

C. Evaluation Metrics

Different image reconstruction methods were compared using the image MSE defined by

$$\text{MSE}(\hat{x}_m) = 10 \log_{10}\left(\|\hat{x}_m - x_m^{\text{true}}\|^2 / \|x_m^{\text{true}}\|^2\right)(\text{dB}), \tag{17}$$

where $\hat{x}_m$ is an image estimate of frame m obtained with one of the reconstruction methods and x_m^{true} denotes the ground truth image. The ensemble bias and standard deviation (SD) of the mean intensity in regions of interest (ROIs) were also calculated to evaluate ROI quantification,

$$\text{Bias} = \frac{1}{c^{\text{true}}}\left|\bar{c} - c^{\text{true}}\right|, \ \text{SD} = \frac{1}{c^{\text{true}}}\sqrt{\frac{1}{N_r - 1}\sum_{i=1}^{N_r} |c_i - \bar{c}|}, \tag{18}$$

where c^{true} is the noise-free intensity and $\bar{c} = \frac{1}{N_r}\sum_{i=1}^{N_r} c_i$ denotes the mean of N_r realizations. c_i is the mean ROI uptake in the ith realization.

D. Demonstration of Attention Map for the Kernel Methods

To understand how the deep kernel method may improve image reconstruction, Fig. 4 illustrates the attention maps for two different "query" pixels, one from the tumor and the other from the white matter region. These attention maps were generated by reshaping the corresponding row of the kernel matrix for a query pixel j. The traditional ML-EM reconstruction can be considered as a special pixel-kernel method for which the kernel matrix is the identity matrix. As illustrated in Fig. 4(b), the attention of ML-EM just focuses on the query pixel j itself. No spatial correlation is explored by this pixel kernel.

The conventional kernel method [3] is able to exploit spatial correlation from pixels that are considered as neighbors of the query pixel j by kNN. The attention is not only on the query pixel but also spreads nonlocally to neighboring pixels ("key") in the whole image. However, these "key" pixels may be falsely identified if k in kNN is large (here $k = 200$) [3]. Without deep learning, the existing kernel model is unable to exclude the effect of those false neighbors. For example, as shown in Fig. 4(c), "key" pixels in the gray-matter and white-matter regions were falsely assigned with high attention for a query pixel from the tumor, and "key" pixels in the gray-matter were also falsely assigned with high attention for a query pixel from the white-matter region.

In comparison, the deep kernel model with training can learn feature extraction from data, which leads to a more appropriate weight to irrelevant "key" pixels even if those pixels are initially included in the k nearest neighbors. Fig. 4(d) shows that with deep learning, attention is predominantly extracted in the tumor region for the tumor query pixel and in the white-matter region for the white-matter query pixel.

E. Image Quality Comparison

Fig. 5 shows the ground-truth activity images and reconstructed images by different reconstruction methods for frame 5 (an early 2-s frame, low count level), frame 15 (a middle 2-s frame, low count level) and frame 55 (a late 1-min frame, relatively high count level), respectively. The results of image MSE in dB are included. The kernel-based methods ((d) and (e)) both achieved a better image quality with lower MSE as compared to the methods without kernel ((b) and (c)). The DIP method [19] suppressed noise well but also resulted in over-smoothness. The proposed deep kernel method achieved a better image quality with lower MSE as compared to other three methods thanks to the improved attention weights embedded in the learned kernel matrix $\boldsymbol{K}$.

Fig. 6(a–c) further show the image MSE plots of frame 5, frame 15 and frame 55 by varying the iteration number in each reconstruction algorithm. For the DIP reconstruction, the results of the first iteration were always better due to the use of four sub-iterations for the tomographic reconstruction step in the ADMM algorithm. The proposed deep kernel method demonstrated a substantial improvement at all later iterations over the conventional kernel method and the DIP method.

The MSE results of all time frames are shown in Fig. 6(d). Here shown are the best MSE (over different iterations) for each frame in different methods. Error bars were calculated over 20 noisy realizations. The DIP method showed an unstable behavior across different frames. In contrast, the deep kernel demonstrated a significant improvement over other methods.

Note that image MSE is only an indicator of global image quality and does not reflect task-specific evaluation. Its weakness is compensated by the ROI quantification results presented in the next subsection.

F.   ROI Quantification Comparison

Fig. 7 shows the trade-off between the absolute bias and SD of different methods for ROI quantification in a blood ROI (Fig. 7a) and a tumor ROI (Fig. 7b and Fig. 7c). The curves were obtained by varying the iteration number from 10 to 200 iterations with an interval of 10 iterations. Note that the uptake in the blood region reached its maximum in frame 5 as shown in Fig. 3b. At a comparable bias level, the proposed deep kernel had a lower noise SD than the conventional kernel method for both the blood ROI and tumor ROI. The results by the DIP method were even worse than the ML-EM results due to over-smoothness, though it had a better image MSE performance as shown in Fig. 6. Some curves in Fig. 7b show a sharp change of direction because the bias at early iterations was negative and became positive due to high noise at late iterations.

G.   Effect of Method Parameters

One parameter that has an important effect on the deep kernel method is the number of training iterations. With increasing iteration number, the training loss was steadily reduced but the corresponding final performance of the PET reconstruction results did not follow this trend. Fig. 8 shows the effect of training iterations on the MSE performance for frame 5, frame 15, and frame 55. The quality of reconstructed PET images may become worse if the training iteration number is too large. This is because the training may start to fit the noise in the composite image prior and the resulting error can be propagated into the trained kernel matrix and final reconstruction. The result here suggests a reasonable choice was 300 iterations, which also worked well for all other frames.

V.   APPLICATION TO PATIENT DATA

A.   Data Acquisition

A cardiac patient scan was performed on the GE Discovery ST PET/CT scanner in 2D mode at the UC Davis Medical Center. The patient received approximately 20 mCi ^{18}F-FDG with a bolus injection, followed by an immediate dynamic scan. The one-hour data are divided into 109 time frames following the schedule $75 \times 2s$, $15 \times 10s$, $10 \times 60s$, and $9 \times 300s$. A low-dose transmission CT scan was performed at the end of PET scan for PET attenuation correction. The projection data size was $249 \times 210 \times 47$ and the image size was $128 \times 128 \times 47$ with a voxel size of $3.91 \times 3.91 \times 3.27$ mm^3. The data correction sinograms of each frame, including normalization, attenuation correction, scattered correction and randoms correction, were extracted using the vendor software and used in the reconstruction process.

B.   Results of Reconstructed PET images

We compared the propsed deep kernel method with the ML-EM, DIP method [19] and conventional kernel method [3]. Because the ADMM algorithm resulted in a very poor DIP reconstruction for this patient dataset, here we instead used the optimization transfer algorithm [32] for the DIP method. Details and advantages of the OT algorithm are described in [32]. The prior images used in the two kernel methods and the DIP method were obtained using four composite images that were reconstructed from four composite frames (one 5-min frame, one 15-min frame and two 20-min frames). Other implementation

settings were as the same as the simulation study. The k in kNN for defining the neighborhood N_j was also set to be 200. All the methods were run for 100 iterations starting from a uniform initial image.

Fig. 9 shows the reconstructed activity images using different algorithms for two early-time high-temporal resolution (HTR) frames (2s/frame), one at $t = 37$s and the other at $t = 145$s. The ML-EM reconstructions were extremely noisy due to the low-count level. The conventional kernel method led to substantial noise reduction but additional noise still remained. Similar to the simulation results, the DIP method successfully suppressed the high noise but also resulted in oversmoothed images and inconclusive separation between the left ventricle and right ventricle. In comparison, the images by the proposed deep kernel method demonstrated a significant improvement with clearer structures and lower noise in the left ventricle cavity and myocardium, though no ground truth is available for the real dataset.

C. Demonstration for Parametric Imaging

Parametric imaging was also performed for the dynamic images of the same subject using a two-tissue compartment model [33]. We used the classic Levenberg–Marquardt algorithm with 50 iterations to solve the optimization problem and the fitting process was implemented using c/c++ programing [34]. For each method, the left ventricle region was used to extract an image-derived input function. Because different reconstruction methods mainly make a difference for early-time frames which have a low count level (Fig. 9), here we focused on parametric imaging of early-dynamic data using the first 150 seconds.

Fig. 10 shows the parametric images of FDG delivery rate K_1. The ML-EM result suffered from heavy noise. The conventional kernel method demonstrated an improvement but still suffered from noise and artifacts. The DIP method largely reduced the noise but also resulted in oversmoothness. It also led to a high K_1 value in the aorta region compared to other three methods. In comparison, the K_1 image obtained by the proposed deep kernel method substantially suppressed the noise and showed a more continuous and clearer myocardium.

Fig. 11 further shows a quantitative comparison of different methods for myocardial ROI quantification in the K_1 image. Here the ROI mean is plotted versus normalized background noise SD by varying the iteration number from 20 to 100 with an interval of 20 iterations. The conventional kernel method outperformed the ML-EM reconstruction noticeably. For a given myocardial K_1 value, the conventional kernel method had a lower liver background noise SD than ML-EM. The DIP method resulted in underestimation of myocardial K_1 compared to other three methods, though the noise was suppressed well. The proposed deep kernel method achieved a better trade-off than all other three methods. For a given ROI mean value (e.g., 0.57), for example, the deep kernel method had the lowest background noise level as compared to the ML-EM and conventional kernel methods. The deep kernel method also had a higher myocardial K_1 value than the DIP method for a given noise level (e.g., 8%) in the liver background. The higher K_1 value was closer to the myocardial ROI mean quantified with the ML-EM reconstruction.

VI.   Discussions

This paper proposed a deep kernel method that learns the trainable components of the neural network model from image prior to enable automated learning of an improved kernel method. Compared to the conventional kernel method [3] that builds the kernel representation using an empirical process, the proposed deep kernel method can learn to extract a more appropriate feature set for building improved kernels from the data, as illustrated in Fig. 4. Compared to the DIP method [19] that introduces a complex non-linear learning in the reconstruction, the deep kernel method only introduces the non-linear learning into the kernel representation, but remains a linear representation for the kernel coefficient image and is therefore easy to be reconstructed from PET data. The comparison results from the simulation and real data studies indicate a better performance of the deep kernel method than other methods.

Similar to the conventional kernel method [3] and the DIP method [19], the proposed method is directly applicable to single subjects, which has been demonstrated for dynamic PET in this paper but can be potentially extended to static image reconstruction if a training pair becomes possible. The prior image must be of relatively high quality. When noise presents, early-stopping can be used to avoid overfitting in the training. Alternatively, regularized training may be explored to address the challenge.

The deep kernel method in this work focused on frame-by-frame image reconstruction in the spatial domain but can be potentially extended to the spatiotemporal domain as used in [6]. The kernel coefficient image $\boldsymbol{\alpha}$ in the deep kernel model can be also further parameterized using a neural network, in a way similar to our other work [32]. In addition, the current study only used the kernel form following the Gaussian function and Euclidean distance. However, it is possible to train an optimized kernel form from the prior data. These modified but more complex methods will be explored in our future work.

Compared to the standard kernel method, the learning of a deep kernel adds an extra computational cost. For the 3D real data study, the training time was 20 minutes as compared to half a minute for the construction of a conventional kernel matrix. However, the extra computational cost may be relatively small when compared to the time (ranging from 30 minutes to several hours) required for the actual kernelized EM reconstruction step (see (8)) for a dynamic PET scan. In addition, the extra time can be further reduced if a large database becomes available to pre-train the optimal kernel construction, for example, using high performance total-body PET scanners (e.g., [35]–[39]), which will also be explored in our future work.

VII.   Conclusion

In this paper, we have developed a new deep kernel method for PET image reconstruction. The proposed deep kernel model allows the construction of kernel representation to be trained from data rather than defined by an empirical process. Computer simulation and patient results have demonstrated the improvement of the deep kernel method over existing methods in dynamic PET imaging.

Acknowledgment

References

[1]. Vaquero J and Kinahan P, " Positron emission tomography: current challenges and opportunities for technological advances in clinical and preclinical imaging systems," Annu. Rev. Biomed. Eng, vol. 17, pp. 385–414, 2015. [PubMed: 26643024]

[2]. Qi J and Leahy RM, "Iterative reconstruction techniques in emission computed tomography," Phys. Med. Biol, vol. 51, no. 15, pp. R541–R578, 2006. [PubMed: 16861768]

[3]. Wang GB and Qi J, "PET image reconstruction using kernel method," IEEE Trans. Med. Imag, vol. 34, no. 1, pp. 61–71, Jan., 2015.

[4]. Hutchcroft W, Wang GB, Chen K, Catana C, and Qi J, "Anatomically-aided PET reconstruction using the kernel method," Phys. Med. Biol, vol. 61, no. 18, pp. 6668–6683, 2016. [PubMed: 27541810]

[5]. Novosad P and Reader AJ, "MR-guided dynamic PET reconstruction with the kernel method and spectral temporal basis functions," Phys. Med. Biol, vol. 61, no. 12, pp. 4624–4645, 2016. [PubMed: 27227517]

[6]. Wang GB, "High temporal-resolution dynamic PET image reconstruction using a new spatiotemporal kernel method," IEEE Trans. Med. Imag, vol. 38, no. 3, pp. 664–674, Mar., 2019.

[7]. Bland J et al. "MR-guided kernel EM reconstruction for reduced dose PET imaging," IEEE Trans. Radiat. Plasma Med. Sci, vol. 2, no. 3, pp. 235–243, May., 2018. [PubMed: 29978142]

[8]. Deidda D et al. "Hybrid PET-MR list-mode kernelized expectation maximization reconstruction," Inverse Probl, vol. 35, no. 4, pp. 044001, 2019.

[9]. Cheng JCK et al. "Dynamic PET image reconstruction utilizing intrinsic data-driven HYPR4D denoising kernel," Med. Phys, vol. 48, no. 5, pp. 2230–2244, 2021. [PubMed: 33533050]

[10]. Reader AJ, Corda G, Mehranian A, Costa-Luis C, Ellis S, and Schnabel JA, "Deep learning for PET image reconstruction," IEEE Trans. Radiat. Plasma Med. Sci, vol. 5, no. 1, pp. 1–24, Jan., 2021.

[11]. Gong K, Berg E, Cherry SR and Qi J, "Machine Learning in PET: From Photon Detection to Quantitative Image Reconstruction," Proc. IEEE Inst. Electr. Electron. Eng, vol. 108, no. 1, pp. 51–68, Jan., 2020.

[12]. Wang T, et al. "Machine learning in quantitative PET: A review of attenuation correction and low-count image reconstruction methods," Phys. Med, vol. 76, pp. 294–306, 2020. [PubMed: 32738777]

[13]. Wang G, Ye J, and De Man B, "Deep learning for tomographic image reconstruction," Nat. Mach. Intell, vol. 2, pp. 737–748, 2020.

[14]. Wang G, Jacob M, Mou X, Shi Y, and Eldar YC, "Deep tomographic image reconstruction: yesterday, today, and tomorrow—editorial for the 2nd special issue "Machine Learning for Image Reconstruction"," IEEE Trans. Med. Imag, vol. 40, no. 11, pp. 2956–2964, Nov., 2021.

[15]. Gong K, Kim K, Cui J, Wu D, and Li Q, "The evolution of image reconstruction in PET: from filtered back-projection to artificial intelligence," PET Clinics, no. 16. vol. 4, pp. 533–542, 2021. [PubMed: 34537129]

[16]. Häggström I, Schmidtlein CR, Campanella G, and Fuchs TJ, "DeepPET: A deep encoder-decoder network for directly solving the PET image reconstruction inverse problem," Med. Imag. Anal, vol. 54, pp. 253–262, May., 2019.

[17]. Lim H, Chun IY, Dewaraja YK, and Fessler JA, "Improved low-count quantitative PET reconstruction with an iterative neural network", IEEE Trans. Med. Imag, vol. 39, no. 11, pp. 3512–3522, Nov., 2020.

[18]. Mehranian A and Reader AJ, "Model-based deep learning PET image reconstruction using forward-backward splitting expectation maximisation," IEEE Trans. Radiat. Plasma Med. Sci, vol. 5, no. 1, pp. 54–64, Jan., 2021.

[19]. Gong K, Catana C, Qi J, and Li Q, "PET image reconstruction using deep image prior," IEEE Trans. Med. Imag, vol. 38, no. 7, pp. 1655–1665, Jul., 2019.

[20]. Gong K, et al. "Iterative PET image reconstruction using convolutional neural network representation," IEEE Trans. Med. Imag, vol. 38, no. 3, pp. 675–685, Mar., 2019.

[21]. Gong K, Catana C, Qi J and Li Q, "Direct reconstruction of linear parametric images from dynamic PET using nonlocal deep image prior," IEEE Trans. Med. Imag, vol. 41, no. 3, pp. 680–689, Mar., 2022.

[22]. Xie Z, et al. "Generative adversarial network based regularized image reconstruction for PET," Phys. Med. Biol, vol. 65, no. 12, pp. 125016, 2020. [PubMed: 32357352]

[23]. Yokota T, Kawai K, Sakata M, Kimura Y, and Hontani H, "Dynamic PET image reconstruction using nonnegative matrix factorization incorporatedwith deep image prior," in Proc. IEEE Int. Conf. Comput. Vis, 2019.

[24]. Xie Z, Li T, Zhang X, Qi W, Asma E, and Qi J, "Anatomically aided PET image reconstruction using deep neural networks," Med Phys, vol. 48, no. 9, pp. 5244–5258, Sep., 2021. [PubMed: 34129690]

[25]. Shepp LA and Vardi Y, "Maximum likelihood reconstruction for emission tomography," IEEE Trans. Med. Imag, vol. MI-1, no. 2, pp. 113–122, Oct., 1982.

[26]. Vaswani A, et al. "Attention is all you need," Proc. Adv. Neural Inf. Process. Syst, pp. 5998–6008, 2017.

[27]. Wang X, Girshick R, Gupta A, and He K, "Non-local neural networks," Proc. IEEE Conf. Comput. Vis. Pattern Recognit., pp. 7794–7803, 2018.

[28]. Li S and Wang GB, "Modified kernel MLAA using autoencoder for PET-enabled dual-energy CT." Philos. Trans. Royal Soc. A, vol. 379, no. 2204, 2021.

[29]. Friedman JH, Bentely J, and Finkel RA, "An algorithm for finding best matches in logarithmic expected time," ACM Trans. Math. Software, vol. 3, no. 3, pp. 209–226, 1977.

[30]. Kramer MA, "Nonlinear principal component analysis using autoassociative neural networks," AIChE J, vol. 37, no. 2, pp. 233–243, 1991.

[31]. He K, Zhang X, Ren S, and Sun J, "Delving deep into rectifiers: Surpassing human-level performance on imagenet classification," Proc. IEEE Int. Conf. Comput. Vis., pp. 1026–1034, 2015.

[32]. Li S, Gong K, Badawi R, Kim E, Qi J, and Wang GB, "Neural KEM: a kernel method with deep coefficient prior for PET image reconstruction," arXiv:2201.01443 [Preprint], Jan 5 2022, doi: available from https://arxiv.org/abs/2201.01443.

[33]. Zuo Y, Badawi R, Foster C, Smith T, López J, and Wang GB, "Multiparametric cardiac 18F-FDG PET in humans: kinetic model selection and identifiability analysis." IEEE Trans. Radiat. Plasma Med. Sci, no. 4, vol. 6, pp. 759–767, 2020. [PubMed: 33778234]

[34]. Wang GB and Qi J, "Generalized algorithms for direct reconstruction of parametric images from dynamic PET data," IEEE Trans. Med. Imag, vol. 28, no. 11, pp. 1717–1726, Nov., 2009.

[35]. Cherry S, Badawi R, Karp J, Moses W, Price P, and Jones T, "Total-body imaging: Transforming the role of PET in translational medicine," Sci. Transl. Med, vol. 9, no. 381, p. eaaf6169, 2017. [PubMed: 28298419]

[36]. Badawi R, et al. "First human imaging studies with the EXPLORER total-body PET scanner," J. Nucl. Med, vol. 60, no. 3, pp. 299–303, 2019. [PubMed: 30733314]

[37]. Spencer B, et al. "Performance evaluation of the uEXPLORER total-body PET/CT scanner based on NEMA NU 2-2018 with additional tests to characterize PET scanners with a long axial field of view," J. Nucl. Med, vol. 62, no.6, pp. 861–870, 2021. [PubMed: 33008932]

[38]. Karp J, et al. "Imaging performance of the PennPET Explorer scanner," J. Nucl. Med, 59: 222–222, 2018.

[39]. Pantel A, et al. "PennPET Explorer: Human imaging on a whole-body imager," J. Nucl. Med, vol. 61, no. 1, pp. 144–151, 2020. [PubMed: 31562224]
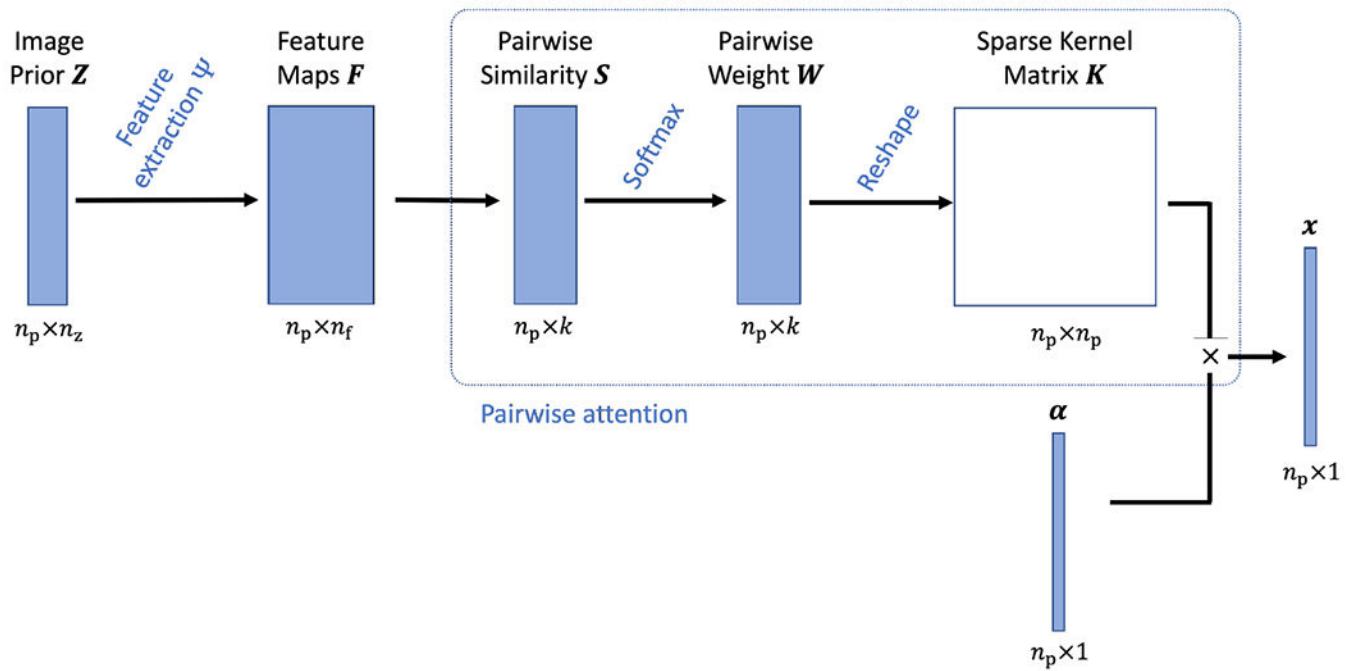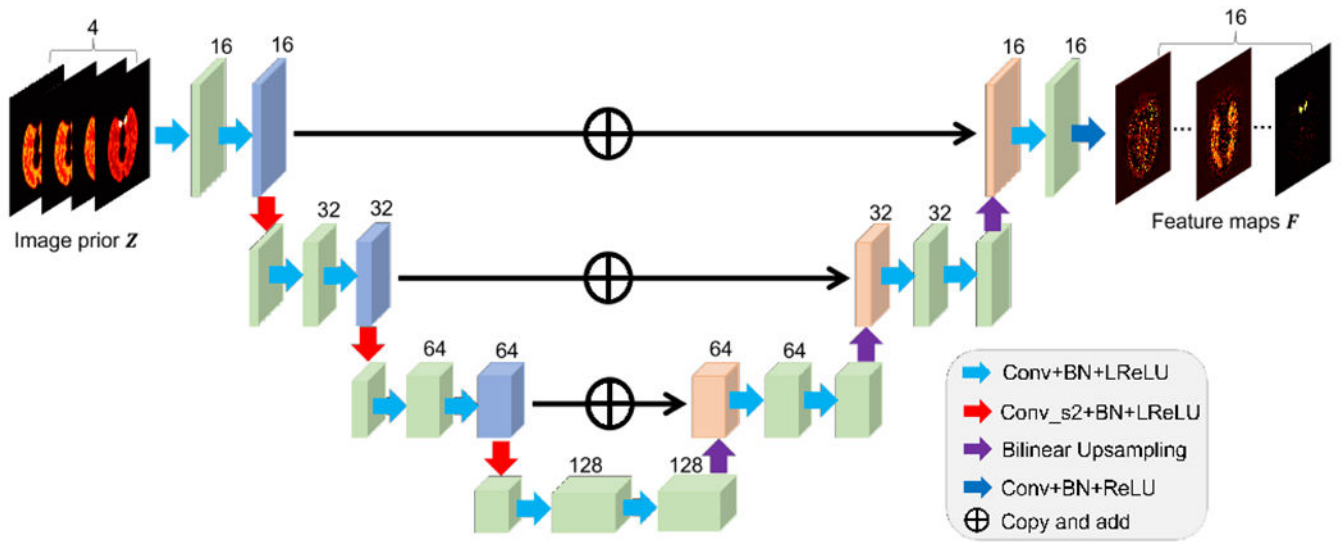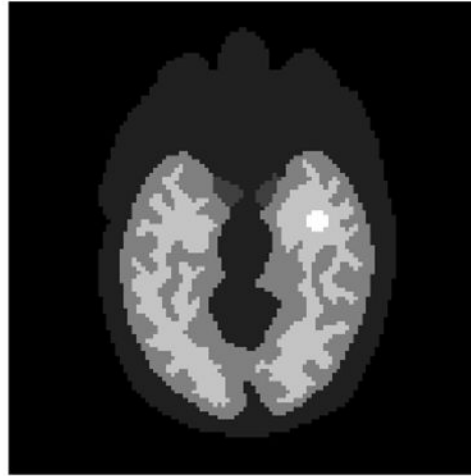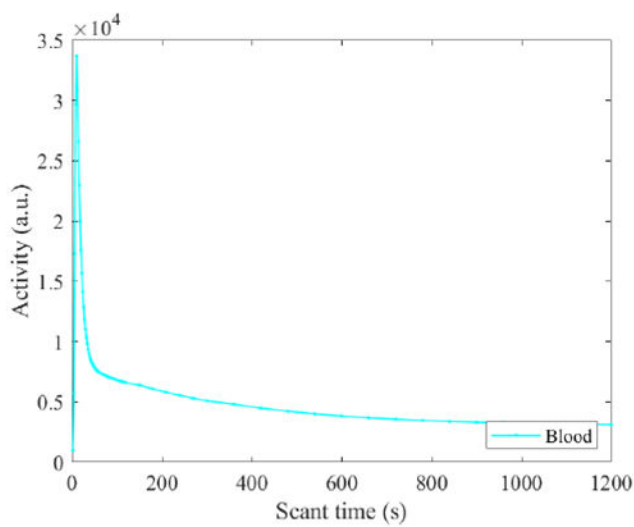
Fig. 1:
The construction of kernel representation for a PET image is described as a series of neural network components. Both the feature extraction and pairwise attention modules are trainable.
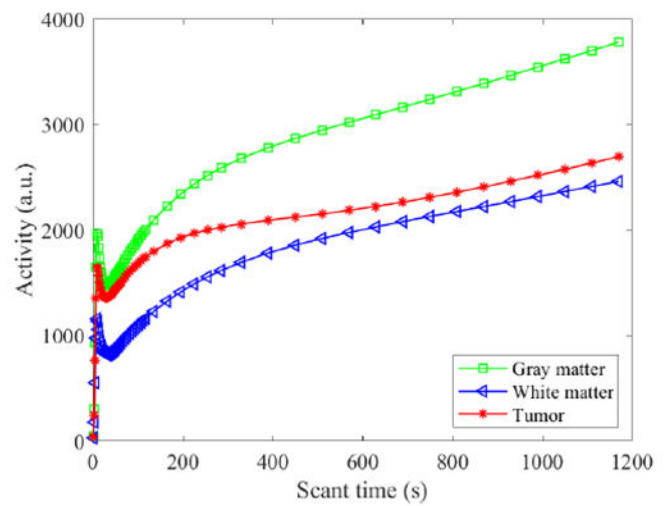
Fig. 2:
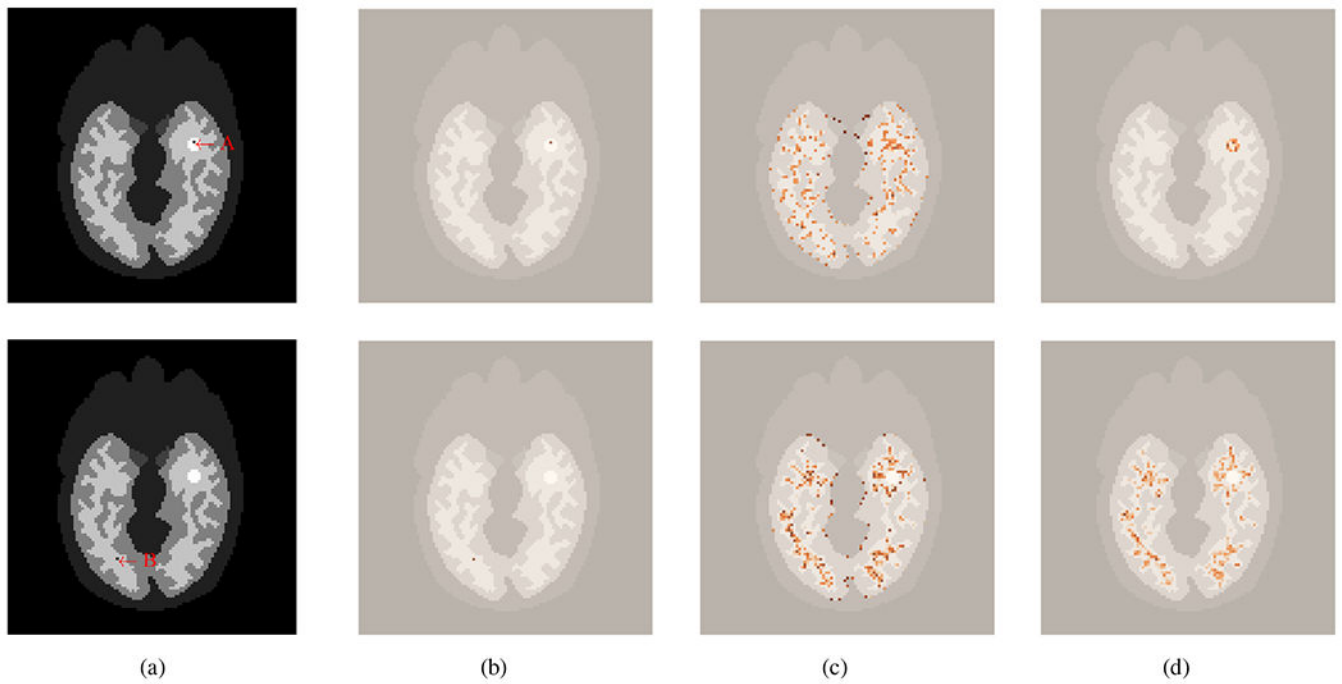Illustration of a residual U-net Ψ used for feature extraction in this work.

(a)

(b)

(c)

Fig. 3:

Digital phantom and time activity curves used in the simulation study. (a) Zubal brain phantom composed of gray matter, white matter, blood pools and a tumor; (b) blood input function; (c) Regional time activity curves of brain regions.

Fig. 4:

Illustration of two query pixels and the attention maps provided by different methods. (a) location of the two query pixels - one in the tumor region (top, A) and the other in the white matter (bottom, B), (b-c) attention maps by traditional ML-EM (b), conventional kernel method (c), and the proposed deep kernel method (d). All the attention maps are overlaid on the structural image.
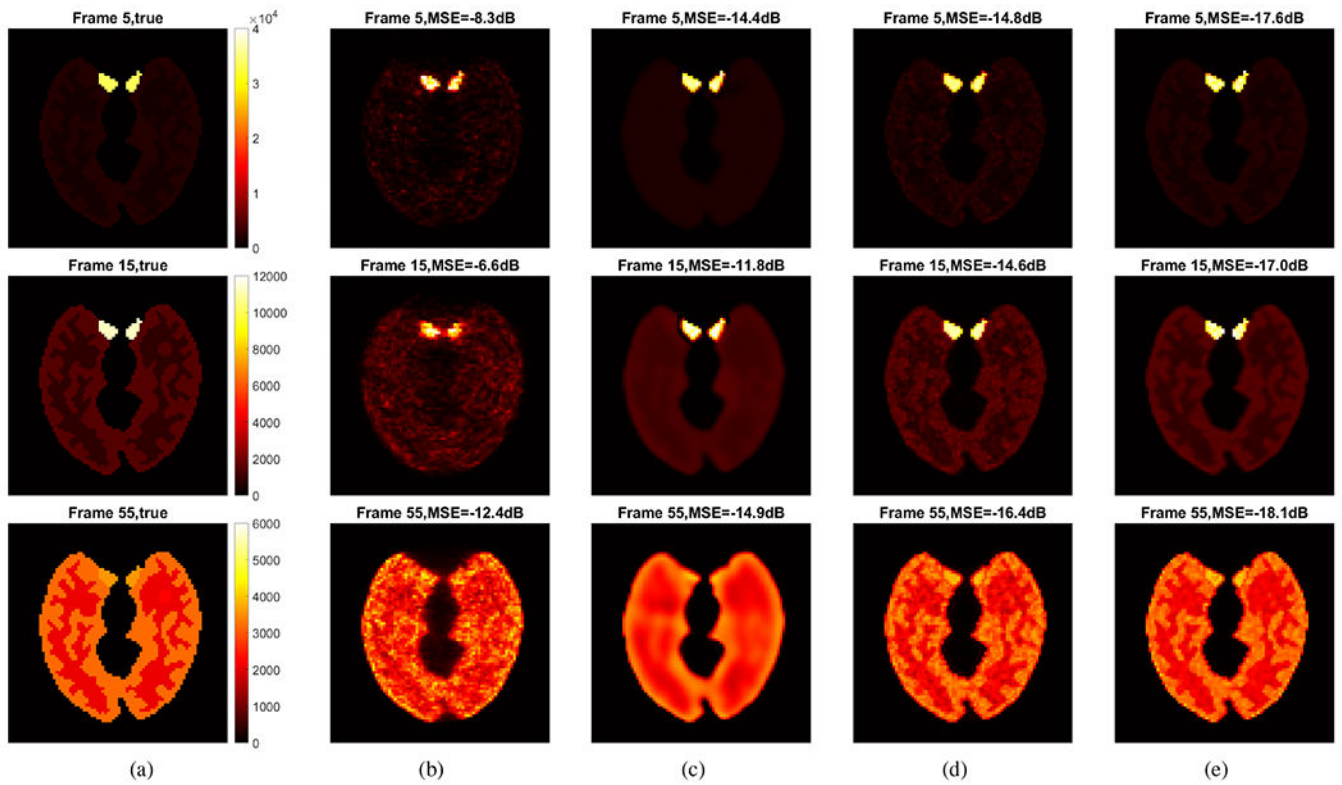
Fig. 5:
True activity images and reconstructed images by different methods for frame 5 (top row), frame 15 (middle row) and frame 55 (bottom row). (a) True images, (b) ML-EM, (c) DIP method, (d) conventional kernel method, and (e) proposed deep kernel method.
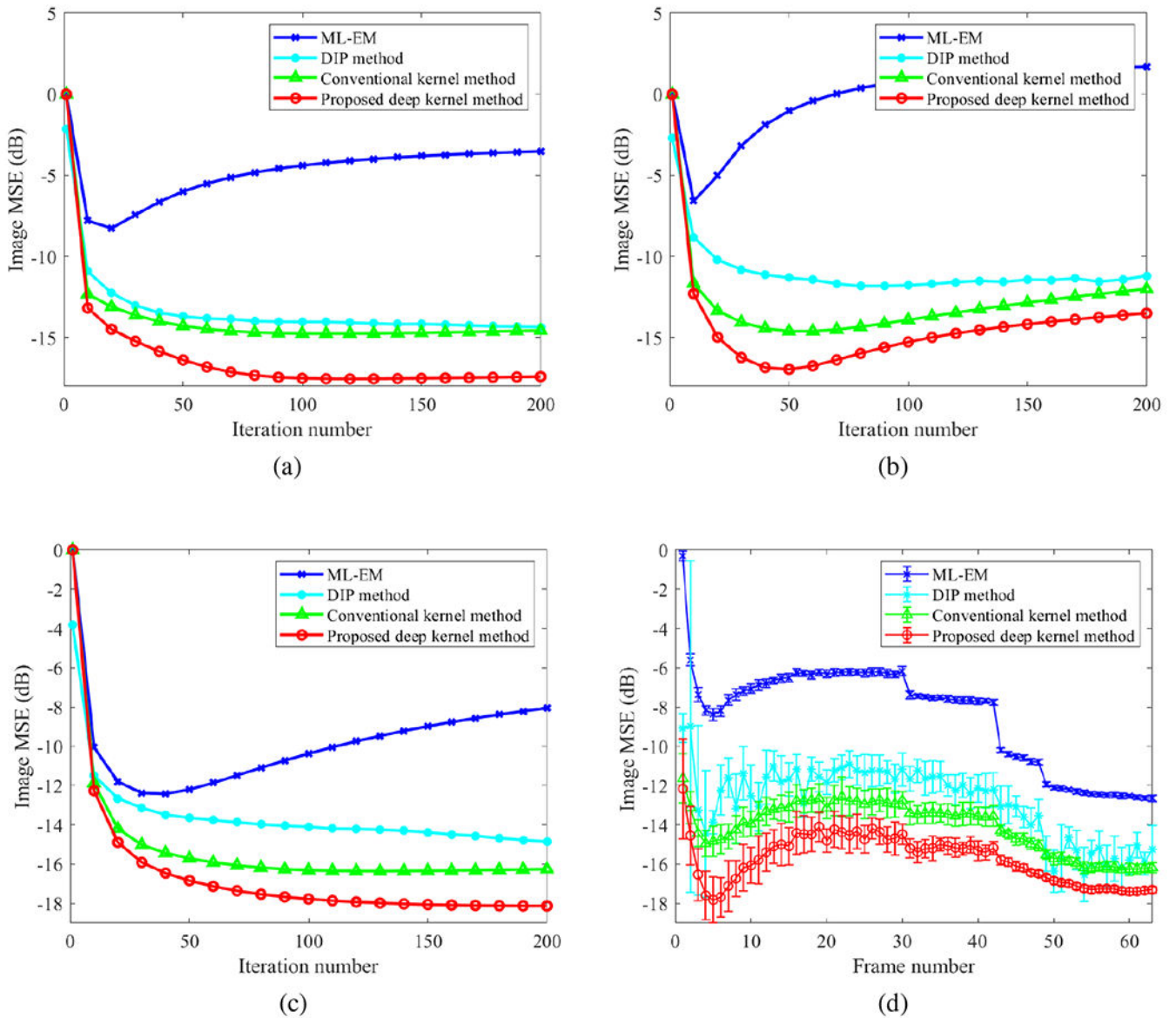
Fig. 6:
Comparison of image MSE for different reconstruction methods. (a-c) plot of image MSE as a function of iteration number for (a) frame 5, (b) frame 15, and (c) frame 55; (d) image MSE of all time frames. The error bars in (d) were obtained from 20 realizations and here the MSE of each frame was minimized over the iteration numbers in different methods.
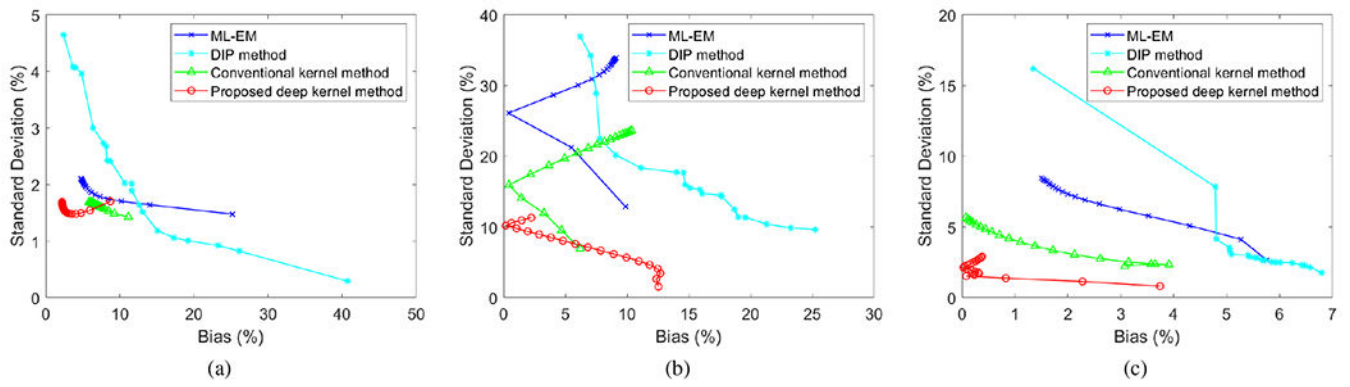
Fig. 7:
Plots of bias-SD trade-off for ROI quantification by varying the iteration number from 10 to 200 with 10 intervals (i.e., from rightmost to leftmost on each curve). (a) Blood ROI in frame 5, (b) tumor ROI in frame 5, (c) tumor ROI in frame 55.
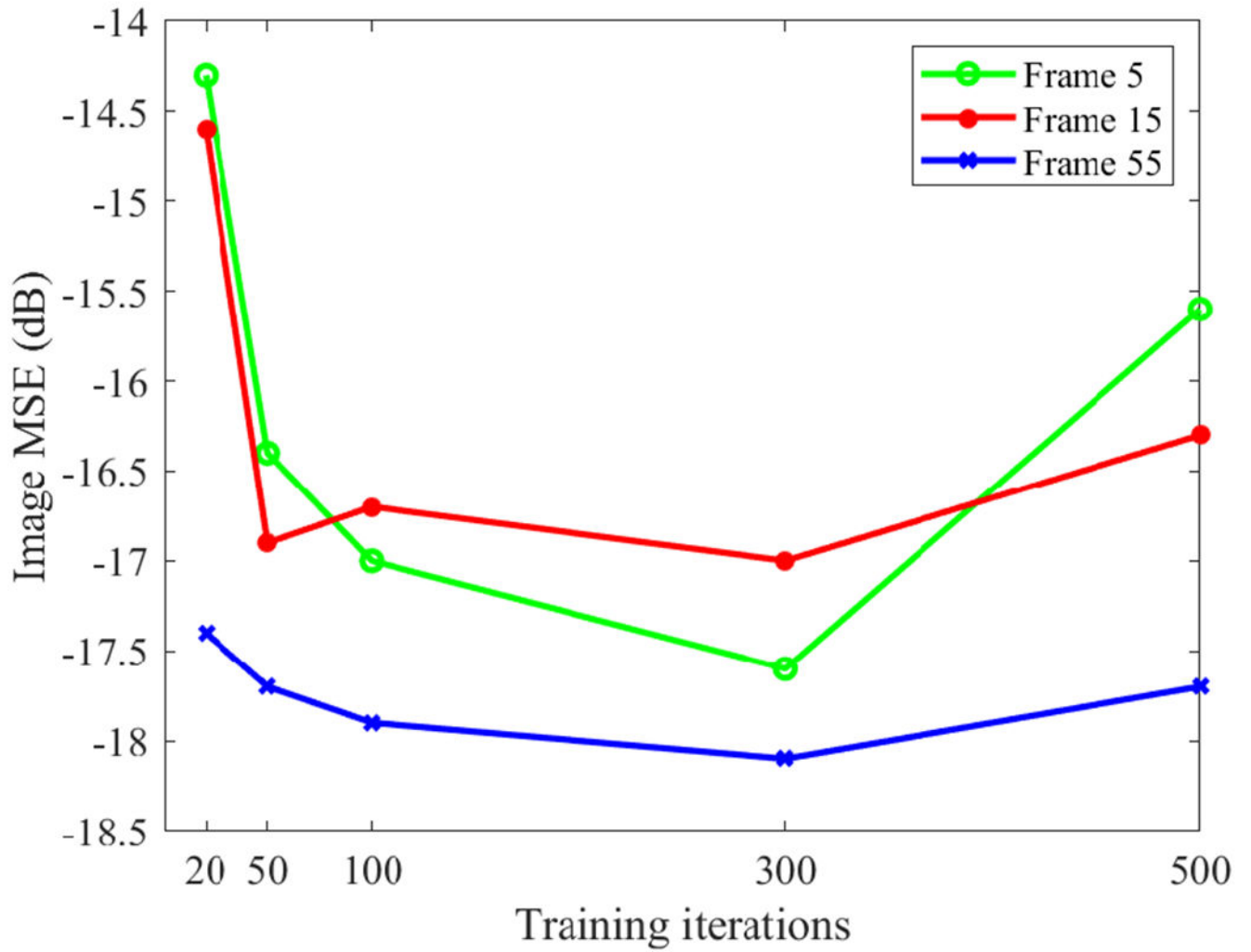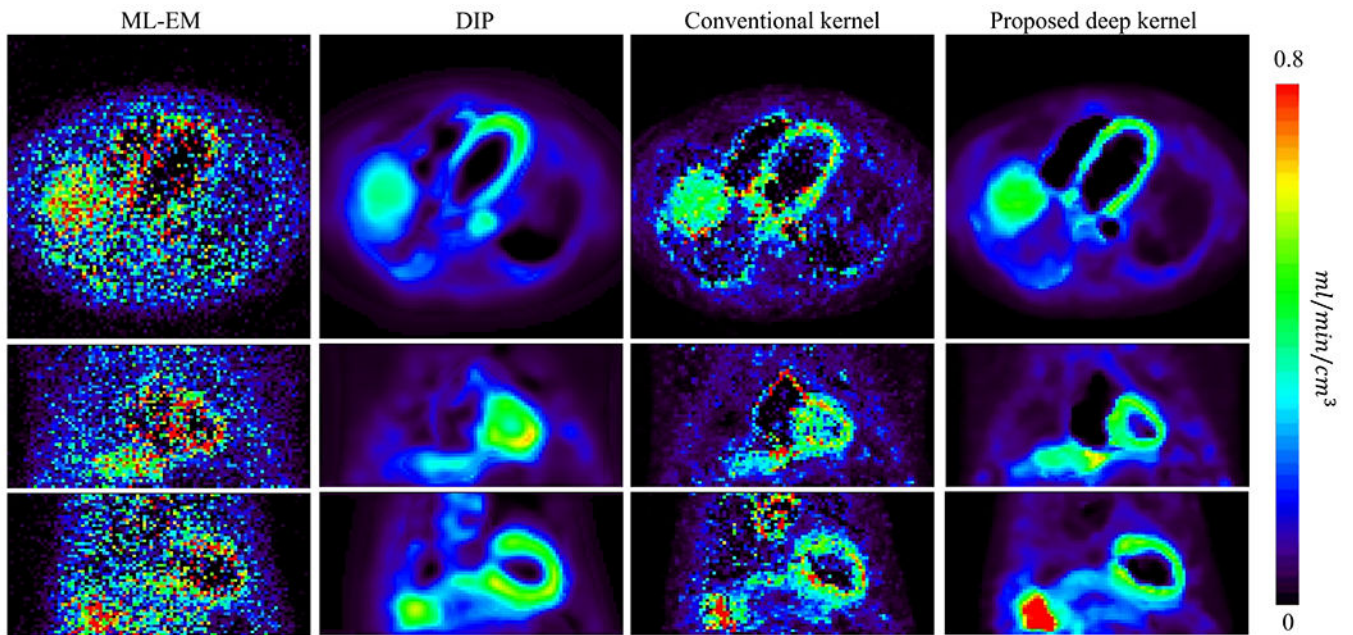
Fig. 8:
Effect of training iterations on the MSE performance of the proposed deep kernel for three different time frames.

(a) Frame 19 at $t = 36 - 38s$

(b) Frame 73 at $t = 144 - 146s$
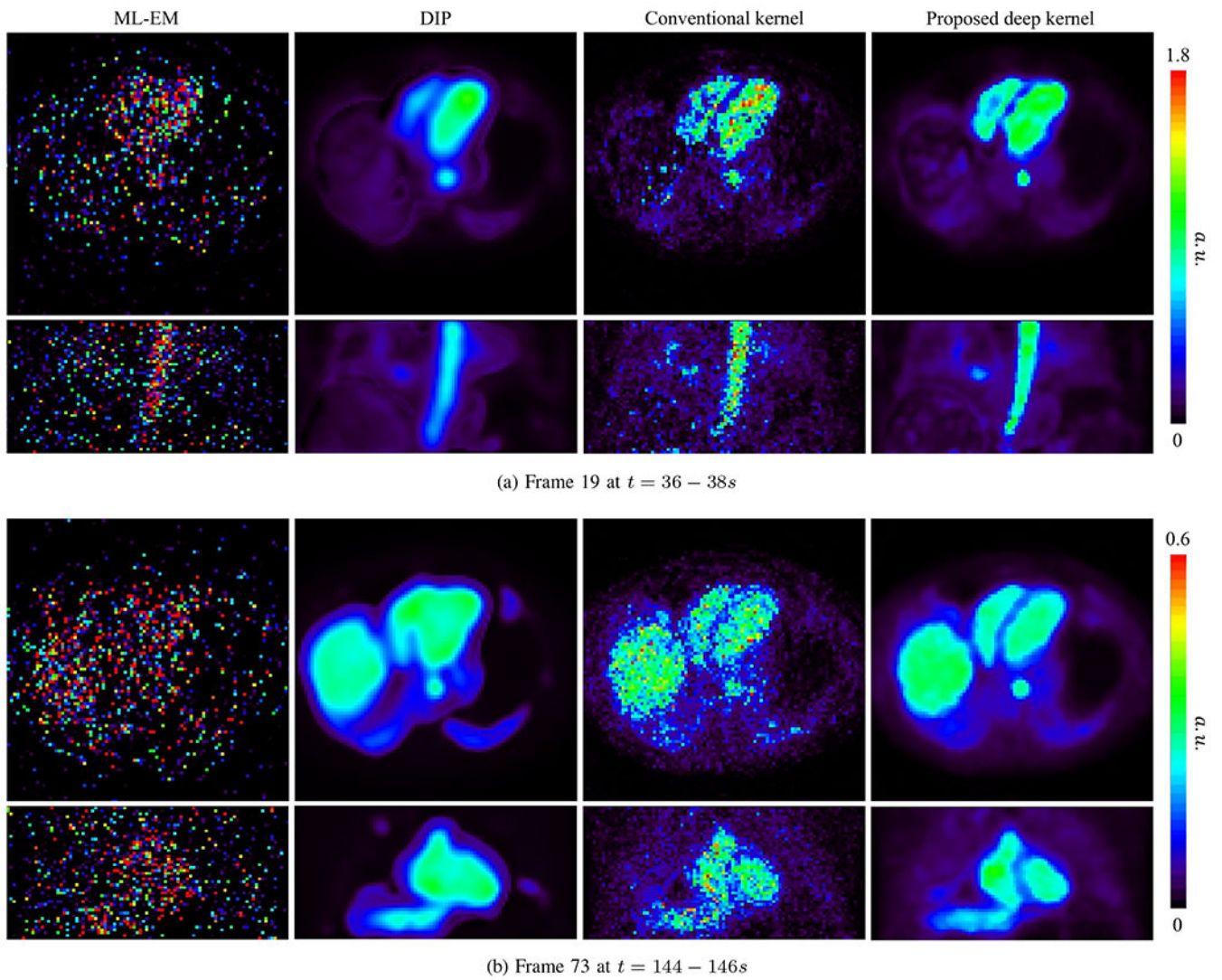
Fig. 9:

Reconstruction of high-temporal resolution frames (2s/frame) at (a) $t = 36 - 38s$ and (b) $t = 144 - 146s$ by different methods: ML-EM, DIP method, conventional kernel method and proposed deep kernel method.

Fig. 10:

Parametric images of K_1 generated from the early-dynamic images reconstructed using ML-EM, DIP method, conventional kernel method and proposed deep kernel method. Each image is shown in transverse, coronal and sagittal views.
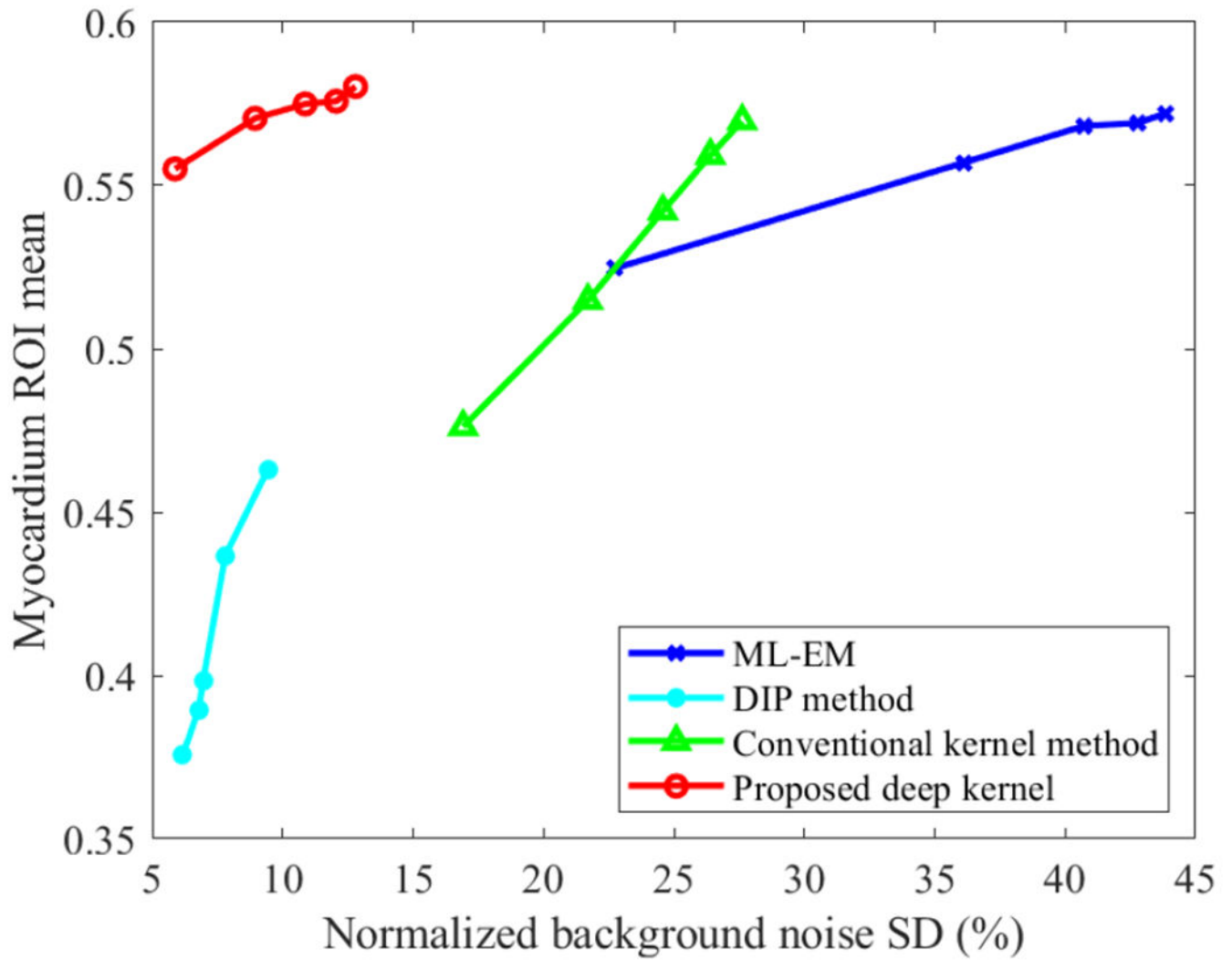
Fig. 11:
Plots of ROI mean of myocardial K_1 versus liver background noise by varying the reconstruction iteration number from 20 to 100 in each method.