

Received March 12, 2021, accepted March 23, 2021, date of publication March 29, 2021, date of current version April 9, 2021.

Digital Object Identifier 10.1109/ACCESS.2021.3069236

Dynamic PET Image Denoising Using Deep Image Prior Combined With Regularization by Denoising

HAO SUN¹, LIHONG PENG, HONGYAN ZHANG¹, YURU HE¹,
SHUANGLIANG CAO¹, AND LIJUN LU¹

Guangdong Provincial Key Laboratory of Medical Image Processing, School of Biomedical Engineering, Southern Medical University, Guangzhou 510515, China

Corresponding author: Lijun Lu (ljlubme@gmail.com)

This work was supported in part by the National Natural Science Foundation of China under Grant 81871437, in part by the Guangdong Basic and Applied Basic Research Foundation under Grant 2019A1515011104 and Grant 2020A1515110683, and in part by the China Postdoctoral Science Foundation Funded Project under Grant 2020M682792. The work of Lijun Lu was supported by the Guangdong Province Universities and Colleges Pearl River Scholar Funded Scheme, in 2018.

ABSTRACT The quantitative accuracy of positron emission tomography (PET) is affected by several factors, including the intrinsic resolution of the imaging system and inherently noisy data, which result in a low signal-to-noise ratio (SNR) of PET image. To address this problem, in this paper, we proposed a novel deep learning denoising framework aiming to enhance the quantitative accuracy of dynamic PET images via introduction of deep image prior (DIP) combined with Regularization by Denoising (RED), as such the method is labeled as DeepRED denoising. The network structure is based on encoder-decoder architecture and uses skip connections to combine hierarchical features to generate the estimated image. The network input can be random noise or other prior images (such as the patient's own static PET image), avoiding the need of high quality noiseless images, which is limited in PET clinical practice due to high radiation dose. Based on simulated data and real patient data, the quantitative performance of the proposed method was compared with conventional Gaussian filtering (GF), non-local mean (NLM), block-matching and 3D filtering (BM3D), DIP and stochastic gradient Langevin dynamics (SGLD) method. Overall, the proposed method can outperform other conventional methods in substantial visual as well as quantitative accuracy improvements (in terms of noise versus bias performance) with and without prior images.

INDEX TERMS Positron emission tomography, deep neural networks, deep image prior, regularization by denoising.

I. INTRODUCTION

Positron emission tomography (PET) is a nuclear medical imaging modality enabling quantitative measurements of biomolecular mechanisms by radioactive tracers in vivo. However, the quantitative accuracy of PET is affected by several factors, including the intrinsic resolution of the imaging system and inherently noisy data, which result in low signal-to-noise ratio (SNR) of PET image [1]. Therefore, it is essential to improve the quantitative accuracy of PET images.

Various pre and post-reconstruction algorithms have been developed by exploiting local statistics, spatiotemporal correlation, or prior anatomical information for PET image

enhancement [2]–[4]. The pre-processing method is carried out in PET sinogram to improve the estimation of physical parameters by using the spatiotemporal correlation in dynamic PET scans [5], [6]. For post-processing algorithms, conventional Gaussian filtering (GF) has been applied for PET image enhancement in clinic. Other post-processing algorithms, such as non-local mean (NLM) [7], wavelet [8], HYPR processing [9], guided image filtering [10], bilateral filtering [11] and kinetics-induced block matching and 5D filtering (KIBM5D) [12] have been proposed, and outperform conventional Gaussian filtering in terms of reducing image noise as well as preserving more image structure details.

The image restoration process is ill-conditioned due to the limited information obtained by PET image itself. Therefore, some techniques seek to incorporate anatomical priors

The associate editor coordinating the review of this manuscript and approving it for publication was Gerardo Di Martino¹.

into image restoration [13], [14]. One strategy is to rely on the explicit boundary or regional information derived from anatomical images by segmentation, which is sensitive to the accuracy of segmentation [15]–[17]. Another strategy not requiring segmentation has also been developed, which is called segmentation-free. These methods based on segmentation-free have the advantage that they are insensitive to potential errors in the segmentation of the anatomical images [18]–[24].

In recent years, deep neural networks have shown great potential in image science, such as image restoration [25], segmentation [26], object detection [27], and image analysis [28], which show better performance than conventional state-of-the-art methods when large amounts of data sets are available. Recently, it has been gradually applied to medical imaging using convolutional neural networks (CNNs) [29]–[33] or generative adversarial networks (GANs) [34]–[36] to improve SNR and reduce artifacts. For medical imaging tasks, it is usually essential to acquire high quality noiseless training image datasets, which are derived from high dose or long-scanned duration as the training labels. However, in the context of positron emission tomography and computed tomography (CT) imaging, high dose CT and PET are limited in routine clinical practice. In the context of MR imaging, the long-scanned duration will improve the SNR of MR images at the cost of increasing movement artifact, especially for cardiac and abdomen imaging. Due to the limitations of data acquisition, available data sets may not be significant or extensive for network training, which will make the network easy to overfit. What's more, collecting data sets and pre-training networks need adequate equipment support and cost much time. Therefore, it is significant to find some effective deep learning methods in case of no training pairs or pre-training.

Recently, the deep image prior (DIP) [37] shows that the structure of a generator network is sufficient to capture a great deal of low-level image statistics prior to any learning without pre-training. A randomly-initialized neural network with random noise input can be used as a handcrafted prior to generate great denoised images, which is a solution in the absence of large data sets. In fact, Gong *et al.* [38] first applied DIP to static PET image reconstruction. Then, Yokota *et al.* [39] proposed a method that reconstructs dynamic PET images from given sinograms by using non-negative matrix factorization incorporated with DIP for appropriately constraining the spatial patterns of resultant images. Furthermore, Cui *et al.* [40] utilized the deep image prior for unsupervised PET image denoising, Hashimoto *et al.* [41] applied DIP in CNN for dynamic brain PET image denoising, and proposed a 4D DIP framework for dynamic PET image denoising [42]. However, these methods are limited to the drawback that the DIP method is easy to overfit (e.g., the update of network needs early stopping to avoid rapid deterioration of image quality). Though the effectiveness of DIP has been proved, its results weaker a little than state-of-the-art alternatives. Fortunately, Cheng *et al.* [43] proposed a novel Bayesian

view of the DIP method, and conducted posterior inference using stochastic gradient Langevin dynamics (SGLD) [44] to avoid the need for early stopping and improve results for denoising task. Mataev *et al.* [45] proposed to bring-in the concept of Regularization by Denoising (RED) as an explicit prior for DIP, which enriches the overall regularization for better performance.

In this paper, we develop a novel deep learning denoising framework aiming to enhance the quantitative accuracy of dynamic PET images via introduction of deep image prior combined with RED. Inspired by [38], [40]–[42], the network input is not only random noise but other prior images. In this work, the static PET image is used as the network input. The use of static image as prior information has been involved in some works. Chen *et al.* [46] proposed a prior image constrained compressed sensing (PICCS) method, which utilized a prior image reconstructed from the union of interleaved dynamical data sets to constrain the CS image reconstruction for the individual time frames. Hashimoto *et al.* [47] utilized a normalized static PET sinogram as the guidance image for PET image noise reduction, which is referred to sinogram-based dynamic image guided filtering (SDIGF) algorithm.

In our denoising framework, the original image (e.g., noisy image) is set as a target image, and mean squared error (MSE) is used to measure the difference between the label and the network output. The network structure is based on encoder-decoder architecture and uses skip connections to combine hierarchical features for generating the estimated image. The solution is formulated and sought by the alternating direction method of multipliers (ADMM) algorithm [48] and split into three sub-optimization steps. In order to stabilize network training, the intensity of the network input is normalized to [0, 1], and the learning rate decays with iterations, which has not been implemented in [38], [40], [41]. Both the computer simulation and real patient data are implemented. The main contributions of this study are as follows: (i) applying DeepRED method to PET image denoising for better quantitative accuracy of dynamic PET image; (ii) providing a solution to overcome the defect that DIP overfits easily in PET imaging; and (iii) comparing the performances of DIP-based method with and without prior images. Overall, The DeepRED method can outperform conventional baseline methods in visual as well as quantitative improvements.

This paper is organized as follows. Section II describes the related works. Section III introduces the theory, optimization process and implementation details. Section IV describes the computer simulation and real patient data used in this work, following specific experiment steps. Section V shows experimental results, while section VI discusses several issues. Finally, conclusions are drawn in Section VII.

II. RELATED WORKS

A. DEEP IMAGE PRIOR

In [37], Dmitry Ulyanov *et al.* proposed that a randomly initialized neural network can be used as a prior with excellent results in standard inverse problems. Given a corrupted target

image, a randomly initialized network with random noise input can approximately restore the uncorrupted image corresponding to the corrupted target image before fitting the target image completely, just using the network structure itself as a prior. Assuming \mathbf{x}_0 is the corrupted image, DIP suggests to solve

$$\theta^* = \arg \min_{\theta} \|\mathbf{x}_0 - T(\theta|z)\|_2^2, \quad \hat{\mathbf{x}} = T(\theta^*|z) \quad (1)$$

where T denotes the neural network, θ^* denotes the network's parameters to be learned, $\hat{\mathbf{x}}$ is the corrected uncorrupted image and z is random noise used as the network input. The training is going without conventional training data pairs.

B. A BAYESIAN PERSPECTIVE ON THE DEEP IMAGE PRIOR

Cheng *et al.* [43] proposed that the DIP is asymptotically equivalent to a stationary Gaussian process prior in the limit as the number of channels in each layer of the network goes to infinity, and derive the corresponding kernel. Conducting posterior inference using stochastic gradient Langevin dynamics can avoid the need of early stopping and improve results for denoising and inpainting tasks. According to (1), the inference procedure can be interpreted as a maximum likelihood estimate (MLE) under a Gaussian noise model: $\hat{\mathbf{x}} = \mathbf{x} + \varepsilon$, $\varepsilon = N(0, \sigma_n^2)$. By adding a suitable prior $\rho(z, \theta)$ over the parameters, the desired image can be derived by $\mathbf{x}^* = \int \rho(z, \theta|\hat{\mathbf{x}})T(z, \theta)dzd\theta$. The integral is replaced by a sample average of a Markov chain that converges to the true posterior. SGLD provides a general framework to derive an MCMC [49] sampler from stochastic gradient descent (SGD) by injecting Gaussian noise to the gradient updates. Let $\omega = (z, \theta)$, the SGLD update follows:

$$\begin{aligned} \Delta\omega &= \frac{\varepsilon}{2} (\nabla_{\omega} \log \rho(\hat{\mathbf{x}}|\omega) + \nabla_{\omega} \log \rho(\omega)) + \eta_t, \\ \eta_t &\sim N(0, \varepsilon) \end{aligned} \quad (2)$$

where ε is the step size and should satisfy the property as follows:

$$\sum_{t=1}^{\infty} \varepsilon_t = \infty, \quad \sum_{t=1}^{\infty} \varepsilon_t^2 < \infty \quad (3)$$

The gradient step sizes and the variances of the injected noise are balanced so that the variance of the samples matches that of the posterior. In experiments, the authors added Gaussian noise to the gradients at each step to estimate the posterior sample averages after a 'burn in' phase.

C. DEEPRED: DEEP IMAGE PRIOR POWERED BY RED

In [45], the authors brought in the concept of RED to DIP for better performance in inverse problems. RED [50] uses a denoising engine to define the regularization term, which can incorporate any image denoising algorithm and guarantee to converge to globally optimal result. RED regularization term is defined as:

$$\rho(\mathbf{x}) = \frac{1}{2} \mathbf{x}^T (\mathbf{x} - f(\mathbf{x})) \quad (4)$$

where \mathbf{x} is the candidate image, $f(\cdot)$ is a denoising engine of choice to plug in the regularization term, like NLM, etc. For such denoisers, Romano *et al.* [50] claim that (4) obeys the gradient rule $\nabla \rho(\mathbf{x}) = \mathbf{x} - f(\mathbf{x})$.

III. METHODOLOGY

A. BACKGROUND

In the field of inverse problems, the measurement $\mathbf{y} \in \mathbf{R}^M$ is given by $\mathbf{y} = \mathbf{A}\mathbf{x} + \mathbf{n}$, where $\mathbf{A} \in \mathbf{R}^{M \times N}$ is a known linear degradation matrix, $\mathbf{x} \in \mathbf{R}^N$ is the original image, and $\mathbf{n} \in \mathbf{R}^M$ is a noise factor. Denoising is obtained for $\mathbf{A} = \mathbf{I}$, where \mathbf{I} is an identity matrix.

B. PROPOSED PET IMAGE DENOISING FRAMEWORK

In this proposed denoising framework, the unknown image is represented as:

$$\mathbf{x} = T(\theta|z) \quad (5)$$

where T denotes the neural network, θ denotes the unknown neural network's parameters, and z is the neural network input. Thus, the original data model can be rewritten as:

$$\mathbf{y} = \mathbf{A}T(\theta|z) + \mathbf{n} \quad (6)$$

The estimated dynamic PET images \mathbf{x} are obtained as follows:

$$\min_{\mathbf{x}, \theta} \frac{1}{2} \|\mathbf{A}T(\theta|z) - \mathbf{y}\|_2^2, \quad \mathbf{x}^* = T(\theta^*|z) \quad (7)$$

Based on [45], merging DIP and RED, the objective function becomes

$$\begin{aligned} \min_{\mathbf{x}, \theta} \frac{1}{2} \|\mathbf{A}T(\theta|z) - \mathbf{y}\|_2^2 + \frac{\lambda}{2} \mathbf{x}^T (\mathbf{x} - f(\mathbf{x})), \\ \text{s.t. } \mathbf{x} = T(\theta|z) \end{aligned} \quad (8)$$

To define the optimization framework, we use the augmented Lagrangian [51] format for the constrained optimization problem in (8), which finally turns into a penalty problem as follows:

$$\begin{aligned} \min_{\mathbf{x}, \theta} \frac{1}{2} \|\mathbf{A}T(\theta|z) - \mathbf{y}\|_2^2 + \frac{\lambda}{2} \mathbf{x}^T (\mathbf{x} - f(\mathbf{x})) \\ + \frac{\mu}{2} \|\mathbf{x} - T(\theta|z) - \mathbf{u}\|_2^2 \end{aligned} \quad (9)$$

where \mathbf{u} denotes the Lagrange multipliers vector for the set of equality constraints, λ and μ are two hyperparameters to be chosen. The objective function in (9) can be solved by the ADMM algorithm iteratively in three steps:

$$\theta^{n+1} = \arg \min_{\theta} \frac{1}{2} \|\mathbf{A}T(\theta|z) - \mathbf{y}\|_2^2 + \frac{\mu}{2} \|\mathbf{x}^n - T(\theta|z) - \mathbf{u}^n\|_2^2 \quad (10)$$

$$\mathbf{x}^{n+1} = \arg \min_{\mathbf{x}} \frac{\lambda}{2} \mathbf{x}^T (\mathbf{x} - f(\mathbf{x})) + \frac{\mu}{2} \|\mathbf{x} - T(\theta^{n+1}|z) - \mathbf{u}^n\|_2^2 \quad (11)$$

$$\mathbf{u}^{n+1} = \mathbf{u}^n - \mathbf{x}^{n+1} + T(\theta^{n+1}|z) \quad (12)$$

In this work, sub-problem (10) is very close to the optimization in DIP modified by a proximity regularization that

forces $T(\theta|z)$ approach to \mathbf{y} and $\mu(\mathbf{x} - \mathbf{u})$. Thus, it will be solved by steepest gradient descent and back-propagation. Sub-problem (11) can be solved in two ways: (i) fixed-point strategy zeroing the derivative of the above with regard to \mathbf{x} . Because (4) obeys the gradient rule $\nabla \rho(\mathbf{x}) = \mathbf{x} - f(\mathbf{x})$, we can get that

$$\lambda(\mathbf{x} - f(\mathbf{x})) + \mu(\mathbf{x} - T(\theta|z) - \mathbf{u}) = 0 \quad (13)$$

Assigning indices to the above equation, (13) follows

$$\lambda(\mathbf{x}^{n+1} - f(\mathbf{x}^n)) + \mu(\mathbf{x}^{n+1} - T(\theta^{n+1}|z) - \mathbf{u}^n) = 0 \quad (14)$$

Thus, the update of \mathbf{x} follows

$$\mathbf{x}^{n+1} = \frac{1}{\lambda + \mu}(\lambda f(\mathbf{x}^n) + \mu(T(\theta^{n+1}|z) + \mathbf{u}^n)) \quad (15)$$

(ii) steepest-descent. The gradient of the regularization term can be surrogated by denoising residual $\mathbf{x} - f(\mathbf{x})$. The update equation would be

$$\mathbf{x}^{n+1} = \mathbf{x}^n - c \left[\lambda(\mathbf{x}^n - f(\mathbf{x}^n)) + \mu(\mathbf{x}^n - T(\theta^{n+1}|z) - \mathbf{u}^n) \right] \quad (16)$$

where c is a hyperparameter to guarantee a descent. The overall algorithm flowchart is presented in Algorithm 1.

Algorithm 1 Algorithm for the Proposed DeepRED Method

Input: Maximum iteration number Max_iter, the RED regularization strength λ , the ADMM free parameter μ , steepest-descent parameters for updating θ , step-size parameter c and image initialization \mathbf{y} , prior image z

- 1: Initialize $n = 0$, $\mathbf{u}^0 = 0$, $\mathbf{x}^0 = \mathbf{y}$, set θ^0 randomly
- 2: **for** $n = 0$ to Max_iter **do**
- 3: Update θ^{n+1} solve $\arg \min_{\theta} \frac{1}{2} \|AT(\theta|z) - \mathbf{y}\|_2^2 + \frac{\mu}{2} \|\mathbf{x}^n - T(\theta|z) - \mathbf{u}^n\|_2^2$ using steepest descent and back-propagation
- 4: Update \mathbf{x}^{n+1} : choose the fixed point or the steepest-descent method
- 5: $\mathbf{x}^{n+1} = \frac{1}{\lambda + \mu}(\lambda f(\mathbf{x}^n) + \mu(T(\theta^{n+1}|z) + \mathbf{u}^n))$ or
- 6: $\mathbf{x}^{n+1} = \mathbf{x}^n - c \left[\lambda(\mathbf{x}^n - f(\mathbf{x}^n)) + \mu(\mathbf{x}^n - T(\theta^{n+1}|z) - \mathbf{u}^n) \right]$
- 7: Update \mathbf{u}^{n+1} : $\mathbf{u}^{n+1} = \mathbf{u}^n - \mathbf{x}^{n+1} + T(\theta^{n+1}|z)$
- 8: $n = n + 1$
- 9: **End for**
- 10: **return** $\hat{\mathbf{x}} = T(\hat{\theta}|z)$

C. NETWORK STRUCTURE

The neural network structure employed in this work is based on encoder-decoder architecture. The overall network architecture is summarized in Figure 1. The encoding patch consists of a 3×3 convolutional layer with a stride two for downsampling instead of using max pooling, followed by a batch normalization (BN) as well as a leaky rectified linear unit (LReLU), and a 3×3 convolutional layer, followed by the BN and the LReLU. The number of feature channels is

fixed to 128 after each downsampling step and upsampling step. The decoding patch consists of bilinear upsampling, a 3×3 convolutional layer followed by the BN and the LReLU, as well as a 3×3 convolutional layer followed by the BN and the LReLU. Due to the checkboard artifact [52], the bilinear interpolation was used instead of the deconvolution upsampling. Besides, we use skip connections to link the encoder path and decoder path in a concatenation way to reduce the number of training parameters and contain structures of different characteristic scales.

D. IMPLEMENTATION DETAILS

The network was run in a GPU (NVIDIA GTX 1080 Ti with 11 GB of memory). We used the Adam optimizer implemented in PyTorch 1.1 on the Pycharm platform, and the MSE between the label and the network output was used as the loss function. Considering the fact that DIP method is easy to overfit, the quality of network output degrades rapidly after a certain number of iterations. Inspired by the supervised training method, the learning rate used in this work decayed with the number of iterations in order to alleviate overfitting. Typically, we set the initial learning rate $lr = 0.005$ decayed $lr \leftarrow 0.9 * lr$ every 1000 iterations. The network output can avoid degrading suddenly due to the excessive learning rate by decaying learning rate. Furthermore, the network output was averaged with an exponential window to acquire stable results. The decay coefficient of exponential moving average was set to 99% for all DIP-based methods. In other words, the results of the DIP-based methods used 1% of the network's output value plus 99% of the average image. The average image of DeepRED method was initialized by the target noisy PET image, and the average images of DIP and SGLD methods were initialized by the network output in the first iteration.

Compromising the convergence speed and performance of the proposed method, Sub-problem (11) was solved by fixed-point strategy, and the hyperparameters $\lambda = 0.2$ and $\mu = 0.2$ were set in this paper. The details of hyperparameter settings will be described in Section VI.

IV. EXPERIMENTAL SETUP

A. SIMULATION STUDY

To validate the proposed method, we employed a simulated ^{18}F -FDG dynamic PET scanning based on the BrainWeb dataset [53], in which a 3D MR brain digital phantom (voxel dimensions, $1.25 \times 1.25 \times 1.25 \text{ mm}^3$) was used. After linear transformation, image cropping, and partial slices extraction, the MR image array size was converted to $256 \times 256 \times 75$. The PET scanning schedule was consistent with those used in [54], including 24 time frames for 60 min: $4 \times 20 \text{ s}$, $4 \times 40 \text{ s}$, $4 \times 60 \text{ s}$, $4 \times 180 \text{ s}$ and $8 \times 300 \text{ s}$. An attenuation image was generated using the discrete MR image by assigning a value of 0 cm^{-1} in air, 0.146 cm^{-1} in bone and 0.096 cm^{-1} in other tissues. The regional time activity curves shown in Figure 2 were assigned to different brain regions

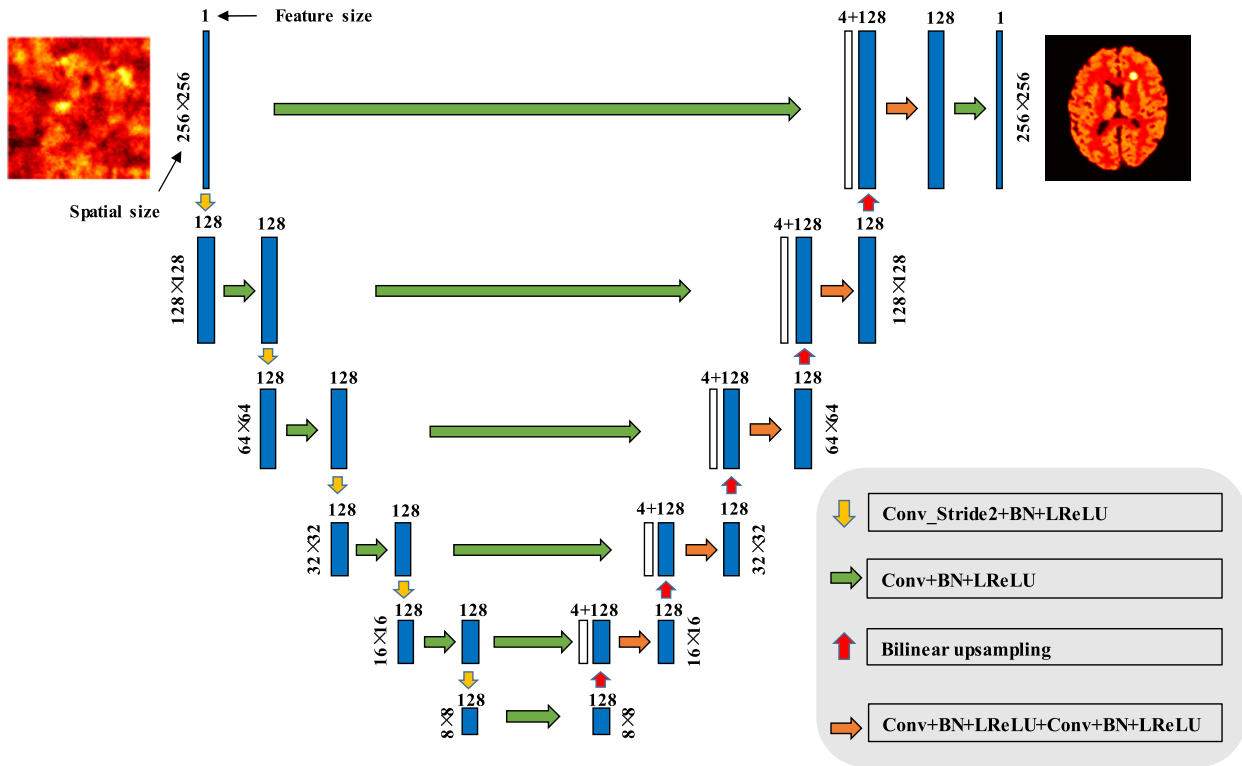


FIGURE 1. The neural network architecture in this work.

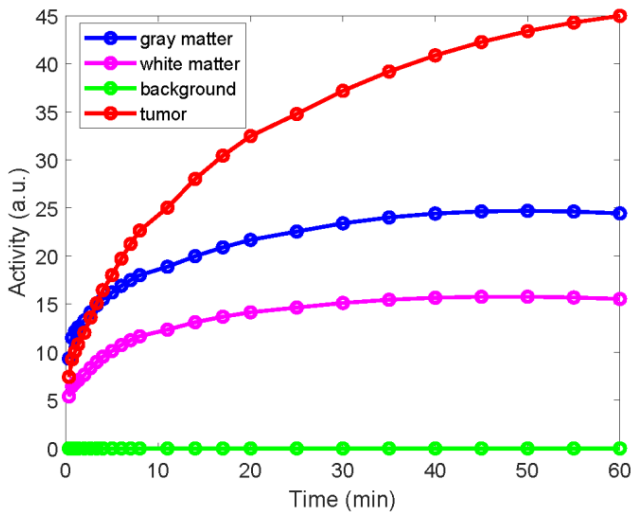


FIGURE 2. The regional time activity curves.

to generate noise-free dynamic PET images. A hot sphere of diameter 20 mm was inserted into the PET image as a tumor region. Frame 24 in dynamic PET images was used as ground truth in this work.

1) TOMOGRAPHY IMAGING

The computer simulation modeled the GE Discovery ST PET/CT scanner, of which the system matrix was shaped by the parallel strip line integration method. Dynamic PET

images were first forward projected using a system matrix to generate noise-free sinogram. After that, the sinogram was convolved with Gaussian filter (full width at half maximum, FWHM = 3.5 voxels) to simulate the partial volume effect. The expected total events were 100M over 60 min. Uniform random events equal to 20% of total counts were added to simulate background events. Then, noisy sinogram data were generated by introducing Poisson distribution. To generate a low-dose noisy image for training, the noisy sinogram data was downsampled to 1/10th of total counts randomly. Finally, the downsampling dynamic frames were reconstructed by Maximum Likelihood Expectation Maximization (MLEM) method with 120 iterations, and frame 24 in downsampling dynamic frames was used as the target noise image to be processed. For the consistency with ground truth, the voxel size of the reconstructed PET image is $1.25 \times 1.25 \times 1.25 \text{ mm}^3$.

We compared the DeepRED method with conventional Gaussian filtering, non-local mean, block-matching and 3D filtering (BM3D) [55], DIP and SGLD method. Here, the FWHM of filter window for GF was set as a chosen parameter from 1mm to 5mm. The radio of search window of NLM was set as a chosen parameter from 3 to 7. The default standard deviation of the additive white Gaussian Noise (AWGN) from 5 to 25 was set as the chosen parameter of BM3D. For the DeepRED method, the NLM denoiser was applied once every 10 iterations. The size of patches of NLM denoiser used in DeepRED was 5×5 and the radio of search window was set to 6.

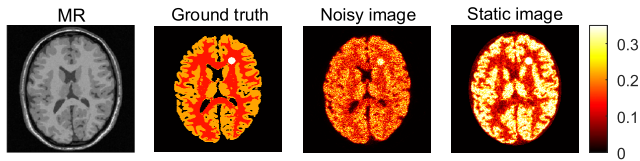


FIGURE 3. A view of simulated brain digital phantom images. These are corresponding to MR brain digital phantom, frame 24 noise-free PET (ground truth), target noisy PET (frame 24) and static PET image from left to right.

As is described in section I, a simulated static PET image is applied as the input of the network to speed up network processing and promote denoising performance. The static PET image was acquired and reconstructed by summing the entire downsampling dynamic PET sinogram data into one static frame. Figure 3 shows the MR brain digital phantom, frame 24 noise-free PET (ground truth), target noise image to be processed (frame 24) and static PET image from left to right.

2) EVALUATION METRIC

For quantitative evaluation, we assessed the quality of denoised PET images by Peak Signal to Noise Ratio (PSNR) and Normalized Root Mean Square Error (NRMSE). Where PSNR is defined as

$$PSNR = 10 \cdot \log_{10} \left[\frac{\max I^2}{\frac{1}{N_R} \sum_{i=1}^{N_R} (I_i - K_i)^2} \right] \quad (17)$$

and NRMSE is defined as

$$NRMSE = \sqrt{\frac{\frac{1}{N_R} \sum_{i=1}^{N_R} (I_i - K_i)^2}{\max I - \min I}} \quad (18)$$

where I and K are the ground truth and the denoised image, respectively. N_R is the number of total pixels in the non-zero ground truth area, $\max I$ and $\min I$ are the maximum and minimum possible pixel value of the ground truth, respectively. Besides, Structural Similarity (SSIM) index was also used. SSIM is defined by

$$SSIM = \frac{(2\mu_I\mu_K + c_1)(2\sigma_{IK} + c_2)}{(\mu_I^2 + \mu_K^2 + c_1)(\sigma_I^2 + \sigma_K^2 + c_2)} \quad (19)$$

where μ is the average value of image, σ^2 is the variance and σ_{IK} is the covariance of I and K . $c_1 = 4 \times 10^{-4}$ and $c_2 = 3.6 \times 10^{-3}$ were used in this work.

In order to ensure the consistency of experimental conditions, the maximum iteration number of all DIP-based methods was preset to 6000, and appropriate denoised PET images were selected based on the highest PSNR during iterations. Due to the overfitting, when the PSNR of images were the same in different iteration numbers, the image with a smaller iteration number was selected as the baseline for comparison. The initial learning rate $lr = 0.005$ was set and decayed $lr \leftarrow 0.9 * lr$ every 1000 iterations. Profile analyses were also implemented for structural comparison with different methods.

On the other hand, the contrast recovery coefficient (CRC) vs. background standard deviation (SD) curves were plotted based on denoised images with different methods for quantitative comparisons. We computed the CRC between the tumor and white matter, and the CRC between the gray matter and white matter. The target region of interests (ROIs) were drawn in both matched gray matter regions and tumor regions. The background ROIs were drawn in white matter regions. For better quantitative evaluation, the CRC was normalized and calculated by

$$CRC = \frac{1}{R} \sum_{r=1}^R \left(\frac{\bar{S}_r}{\bar{B}_r} - 1 \right) / \left(\frac{\bar{S}_{true}}{\bar{B}_{true}} - 1 \right) \quad (20)$$

where R represents the total number of realizations and is set to 10. $\bar{S}_r = 1/N_s \sum_{k=1}^{N_s} S_{r,k}$ denotes the mean activity of the target ROIs in realization r , $\bar{B}_r = 1/N_b \sum_{k=1}^{N_b} B_{r,k}$ denotes the mean activity of the white matter region (background ROIs) in realization r . N_s and N_b are the number of target ROIs and background ROIs in denoised images respectively, where $N_b = 16$, $N_s = 1$ for tumor region and $N_s = 16$ for gray matter regions. The size of gray matter ROIs and background ROIs were all 12×12 pixels. $\bar{S}_{true} = 1/K_s \sum_{m=1}^{K_s} S_{r,m}$ denotes the mean activity of the target region of interest in ground truth images, and $\bar{B}_{true} = 1/K_b \sum_{m=1}^{K_b} S_{b,m}$ denotes the mean activity of the white matter region (background ROIs) in ground truth. K_s and K_b are the number of target ROIs and background ROIs in ground truth respectively, where $K_b = 16$, $K_s = 1$ for tumor region and $K_s = 16$ for gray matter regions. The background SD was calculated by

$$SD = \frac{1}{R} \sum_{r=1}^R \frac{1}{N_b} \sqrt{\sum_{k=1}^{N_b} (S_{r,k} - \bar{B}_r)^2} \quad (21)$$

The network structure for simulation study sets was the same as in Figure 1.

B. REAL DYNAMIC PET BRAIN DATA SETS

Five patient data sets from Alzheimer’s Disease Neuroimaging Initiative (ADNI) [56] were employed in this study. The patients were injected with 50 ± 0.5 mCi (185 MBq) of ^{18}F -FDG injection. PET scanning was performed on a Siemens PET/CT scanner, including 33 time frames for 60 min: 1×10 s, 12×5 s, 2×10 s, 3×30 s, 3×60 s, 2×120 s and 10×300 s, and PET images were reconstructed by the ordered subset expectation maximization (OSEM) method. The PET image matrix size is $128 \times 128 \times 63$ pixels and the voxel size is $2.2 \times 2.2 \times 2.4$ mm³.

For lack of static PET images in the original patient data sets, the voxels of all frames of dynamic PET image were added and averaged in image domain to obtain approximate clinical static PET image for each patient. Then the intensity of the approximate clinical static PET image was scaled to the range $[0, 1]$ for network training. In order to adapt the network structure, the 34th slice of frame 22 PET image was chosen to employ for each patient, and the corresponding

34th slice static PET image was applied as the network input. In this paper, only the results of static PET image input for network training were shown in view of the better performance compared with the result of random noise input in real data set.

We also compared the DeepRED method with conventional Gaussian filtering, non-local mean, block-matching and 3D filtering, DIP and SGLD method. For patient data sets, the FWHM of filter window for GF was set as a chosen parameter from 1 mm to 5 mm. The radio of search window of NLM was set as a chosen parameter from 2 to 6. The default standard deviation of the additive white Gaussian Noise from 5 to 25 was set as the chosen parameter of BM3D. When using the DeepRED method, we applied the NLM denoiser once every 10 iterations. The size of patches of NLM denoiser used in DeepRED was 5×5 and the radio of search window was set to 2. To ensure the consistency of experimental conditions, the maximum iteration number of all DIP-based methods was preset to 6000. The initial learning rate $lr = 0.005$ was set and decayed $lr \leftarrow 0.9 * lr$ every 1000 iterations.

For lack of ground truth, five ROIs in the noisy PET brain image were selected for quantitative comparison. The five ROIs were shown in Figure 8. The uptakes of ROI 1, ROI 3 and ROI 5 increased with the increase of frames, which were the high activity regions. The uptakes of ROI 2 and ROI 4 decreased with the increase of frames, which were the low activity regions and used as background ROIs. The contrast recovery (CR) between the high activity regions with the low activity regions was calculated by

$$CR = \frac{1}{N} \sum_{l=1}^N \frac{|\bar{x}_l - \bar{x}_{ref}|}{\bar{x}_{ref}} \quad (22)$$

where $\bar{x}_l = 1/N_s \sum_{k=1}^{N_s} x_{l,k}$ denotes the mean activity of selected ROIs, $\bar{x}_{ref} = 1/N_b \sum_{m=1}^{N_b} x_{l,m}$ denotes the mean activity of the reference ROIs, and N is the number of realizations, which was set to 10 here. In this work, the target ROIs corresponding to ROI1, ROI3, and ROI5, the reference ROIs corresponding to ROI2 and ROI4. The network structure for the real data sets was the same as in Figure 1.

V. RESULTS

We evaluated the proposed DeepRED algorithm in comparison with conventional denoising methods, including GF, NLM, BM3D, DIP and SGLD algorithms. The parameters in all algorithms were first optimized separately, followed by vertical profile analysis and quantitative comparison in terms of regional contrast recovery coefficient vs. the background standard deviation performance.

A. SIMULATION RESULTS

1) THE NETWORK INPUT

In order to provide a more direct visual impression of the denoised images, Figure 4 shows ground truth, noisy image (low dose), static image and denoised images using different

algorithms: GF, NLM, BM3D, DIP_r, SGLD_r, DeepRED_r, DIP_s, SGLD_s and DeepRED_s methods for one simulated realization, below each image is the zoomed-in ROI of the area indicated by the blue box in the upper corner reference image. Among these methods, DIP_r, SGLD_r and DeepRED_r used random noise as the network input, while DIP_s, SGLD_s and DeepRED_s used static image as the network input. For GF, the FWHM was set to 3 mm. For NLM, the search window was set to 5×5 . For BM3D, the standard deviation of the AWGN was set to 20.

Compared with the GF and NLM post-processing methods, all DIP-based methods and BM3D methods have lower noise and reveal more cortex structures. Compared with BM3D method, all the DIP-based methods can preserve more image details features, especially at gray matter regions, and the DeepRED method performs best. Furthermore, compared with DIP_r, SGLD_r and DeepRED_r methods, DIP_s, SGLD_s and DeepRED_s can retain more image details features. Actually, using static image as the network input can greatly accelerate the speed of network processing, which will be further explained in the following discussion in section VI.

In this work, we took into account the use of decaying learning rate and recorded the mean value and standard deviation of the quantitative results achieved by 10 realizations for different methods. The corresponding quantitative evaluation results are shown in Table 1, and the best value of each metric is marked in bold black. In Table 1, type ND means not decaying learning rate, type D means decaying learning rate and type NA means not applicable. As is shown in Table 1, all DIP-based methods perform better than traditional GF and NLM methods in terms of all indicators. DIP_r and SGLD_r have lower SSIM than BM3D method, while DeepRED_r performs better than BM3D in terms of SSIM. In this work, SGLD_r performs a little weaker than DIP_r, which is somewhat different from the experimental results in [43]. This phenomenon may be due to the fact that the noise in the PET image domain is not a standard Gaussian distribution. Compared with DIP_r, SGLD_r and DeepRED_r methods, DIP_s, SGLD_s and DeepRED_s can acquire great improvements in terms of all indicators, which confirms the effectiveness of using prior image as the network input. For DeepRED_r and DeepRED_s, decaying learning rate can improve the network's denoising performance, while this result is not obvious for the DIP and SGLD algorithms. However, the use of decaying learning rate will alleviate overfitting and avoid network output degrading suddenly. Overall, the DeepREDs method outperforms other methods in terms of all indicators.

2) VERTICAL PROFILE ANALYSIS

To provide an intuitionistic evaluation performance of preserving the edge details, Figure 5 shows the vertical profile analysis of the images shown in Figure 4. For a better demonstration, we divide the results into Figure 5 (a) and Figure 5 (b) based on different network inputs. Figure 5 (a) shows the profiles of the results from ground truth, GF, NLM, BM3D, DIP_r, SGLD_r and DeepRED_r. Figure 5 (b) shows the profiles

TABLE 1. The Mean Value and Standard Deviation of Quantitative Evaluation Results (PSNR: dB, SSIM, and NRMSE) on Brain Digital Phantom with Different Image Denoising Algorithms, Including GF, NLM, BM3D, DIP_r, SGLD_r, DeepRED_r, DIP_s, SGLD_s and DeepRED_s Methods. Type ND Means not Decaying Learning Rate, Type D Means Decaying Learning Rate, and Type NA Means not Applicable.

Method	Type	PSNR(dB)	SSIM	NRMSE
Noisy	NA	27.5322	0.8037	0.0420
GF	NA	29.8750	0.8773	0.0321
NLM	NA	29.9628	0.8837	0.0317
BM3D	NA	30.1565	0.8922	0.0311
DIP _r	ND	30.2281±1.81E-02	0.8902±5.71E-04	0.0308±6.42E-05
	D	30.1818±2.29E-02	0.8860±1.00E-03	0.0310±8.79E-05
SGLD _r	ND	30.1160±2.74E-02	0.8889±1.21E-03	0.0312±9.85E-05
	D	30.1040±2.06E-02	0.8882±1.34E-03	0.0312±7.39E-05
DeepRED _r	ND	30.3618±1.13E-02	0.8956±5.81E-04	0.0303±3.95E-05
	D	30.3671±1.87E-02	0.8954±8.68E-04	0.0303±6.51E-05
DIP _s	ND	30.5929±2.34E-02	0.8977±1.09E-03	0.0295±8.04E-05
	D	30.6067±3.68E-02	0.8890±2.84E-03	0.0295±1.39E-05
SGLD _s	ND	30.6259±2.42E-02	0.9008±1.50E-03	0.0294±8.21E-05
	D	30.6087±1.85E-02	0.9005±1.33E-03	0.0295±1.69E-04
DeepRED _s	ND	30.6862±2.92E-02	0.9019±1.69E-03	0.0292±9.82E-05
	D	30.7134±1.39E-02	0.9026±1.17E-03	0.0291±4.66E-05

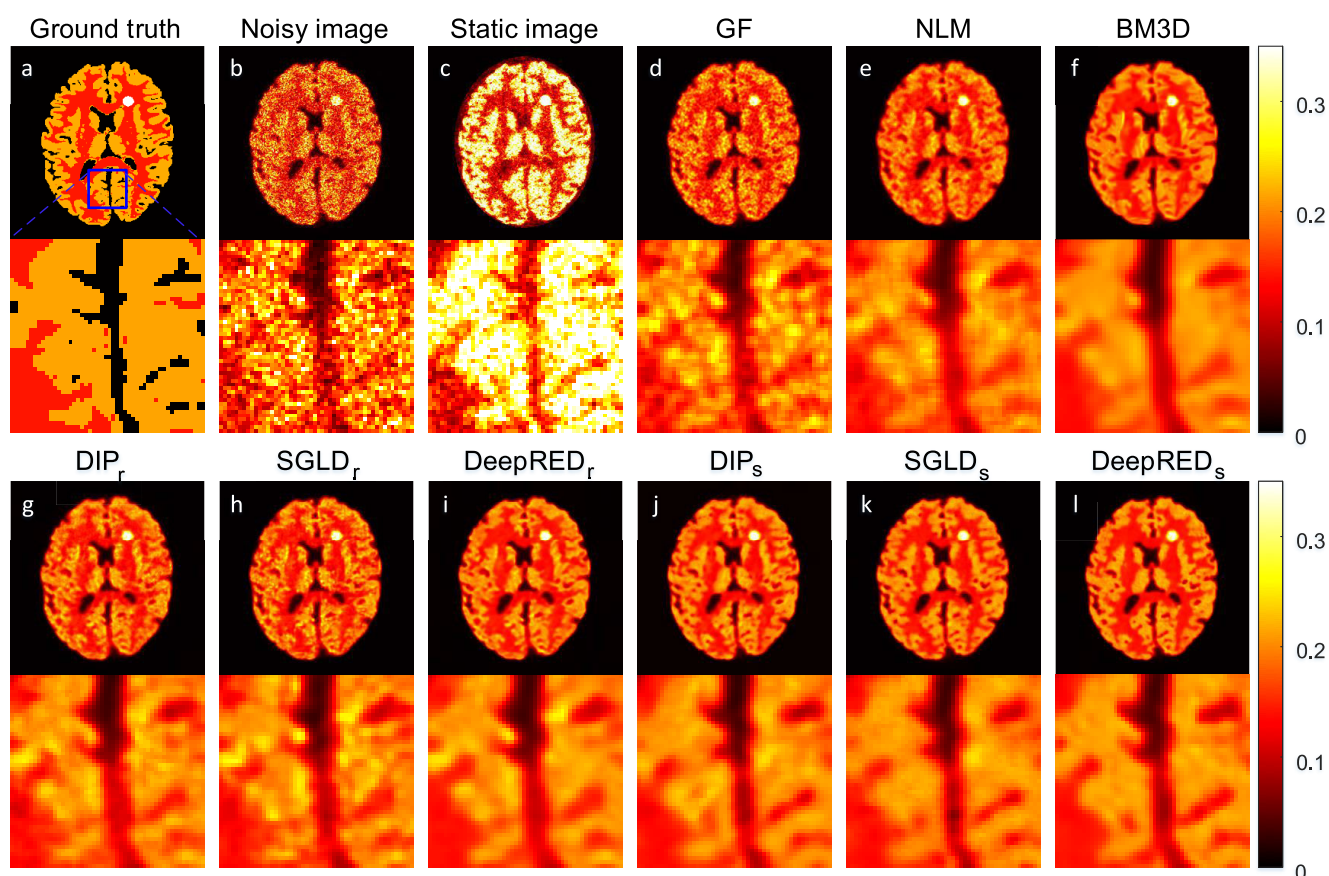


FIGURE 4. A view of (a) Ground truth; (b) noisy image (low dose); (c) static image; (d)-(f) the post-processed PET images using the Gaussian filter with FWHM = 3mm, NLM with 5 × 5 search window size and BM3D with 20 AWGN standard deviation, respectively; (g)-(i) the post-processed PET images using DIP_r, SGLD_r and DeepRED_r methods respectively (Random noise was used as the network input.); (j)-(l) the post-processed PET images using DIP_s, SGLD_s and DeepRED_s methods respectively (Static image was used as the network input). Below each image is the zoomed-in ROI of the area indicated by the blue box in the upper corner reference image.

of the results from ground truth, GF, NLM, BM3D, DIP_s, SGLD_s and DeepRED_s. It can be observed that the profiles of the results from DeepRED_r and DeepRED_s are more closed

to ground truth than those from other methods. Besides, DeepRED_r and DeepRED_s are smoother in the regions with large intensity variation than other methods. These results

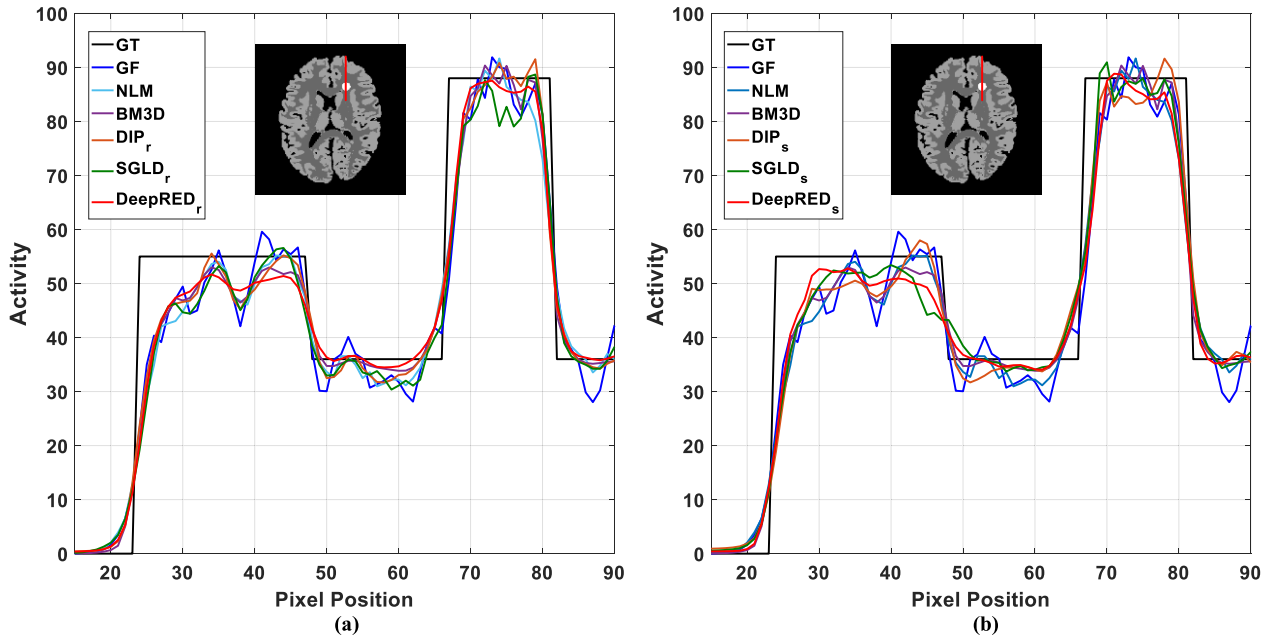


FIGURE 5. The vertical profile analysis including (a) the profiles of the results from ground truth, GF, NLM, BM3D, DIP_r , $SGLD_r$ and $DeepRED_r$, and (b) the profiles of the results from ground truth, GF, NLM, BM3D, DIP_s , $SGLD_s$ and $DeepRED_s$. The vertical profile is located at the pixel positions $x = 155$ and y from 15 to 90, which is shown as the red line on the brain digital phantom.

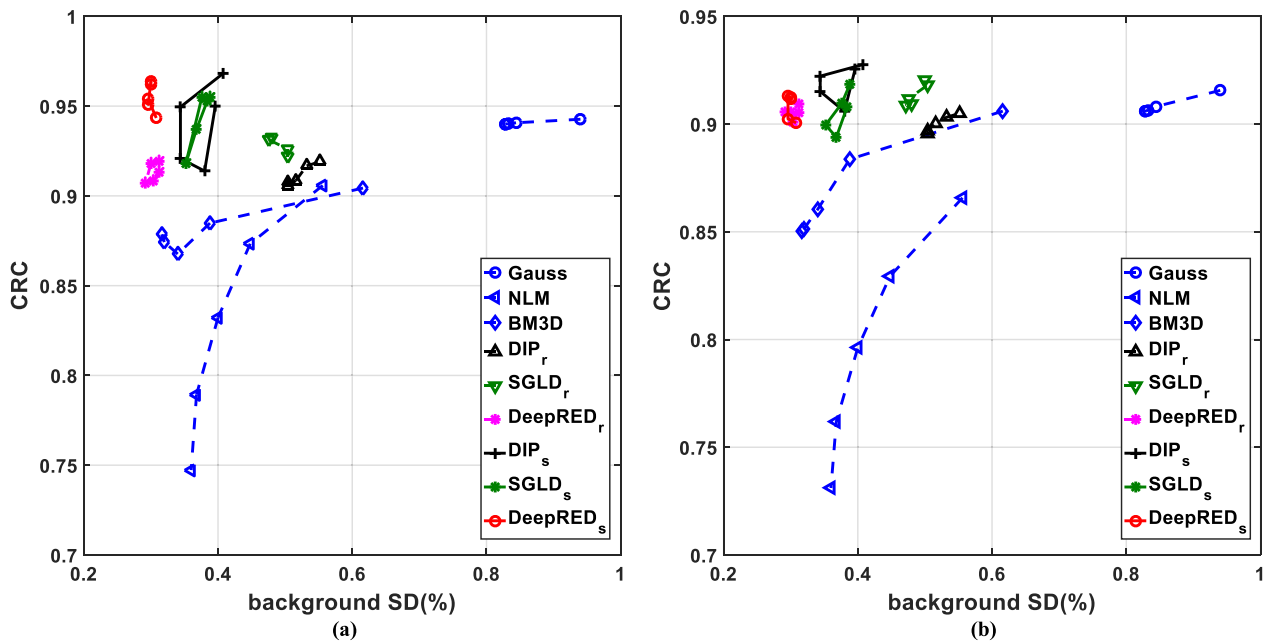


FIGURE 6. Plots of regional CRC vs. background SD curves of estimated images (generated with varying parameters) using different denoising algorithms at (a) the gray matter region and (b) the tumor region.

further verify that the proposed DeepRED algorithm has a better noise reduction capability while preserving edge structures.

3) REGIONAL CRC VS. BACKGROUND SD

To provide a regional contrast recovery evaluation performance of the denoising algorithms, Figure 6 depicts plots of regional CRC vs. background SD curves of estimated images (generated with varying parameters) for gray matter

and tumor region. For GF, the full width at half maximum increased from 1mm to 5mm. For NLM, the search window varied from 3×3 to 7×7 . For BM3D, the standard deviation of the AWGN ranged from 5 to 25. The results from DIP-based methods were generated by varying the iteration numbers.

As is shown in Figure 6, all DIP-based methods outperform traditional methods (including GF, NLM, and BM3D), and the DeepRED method performs better than DIP and

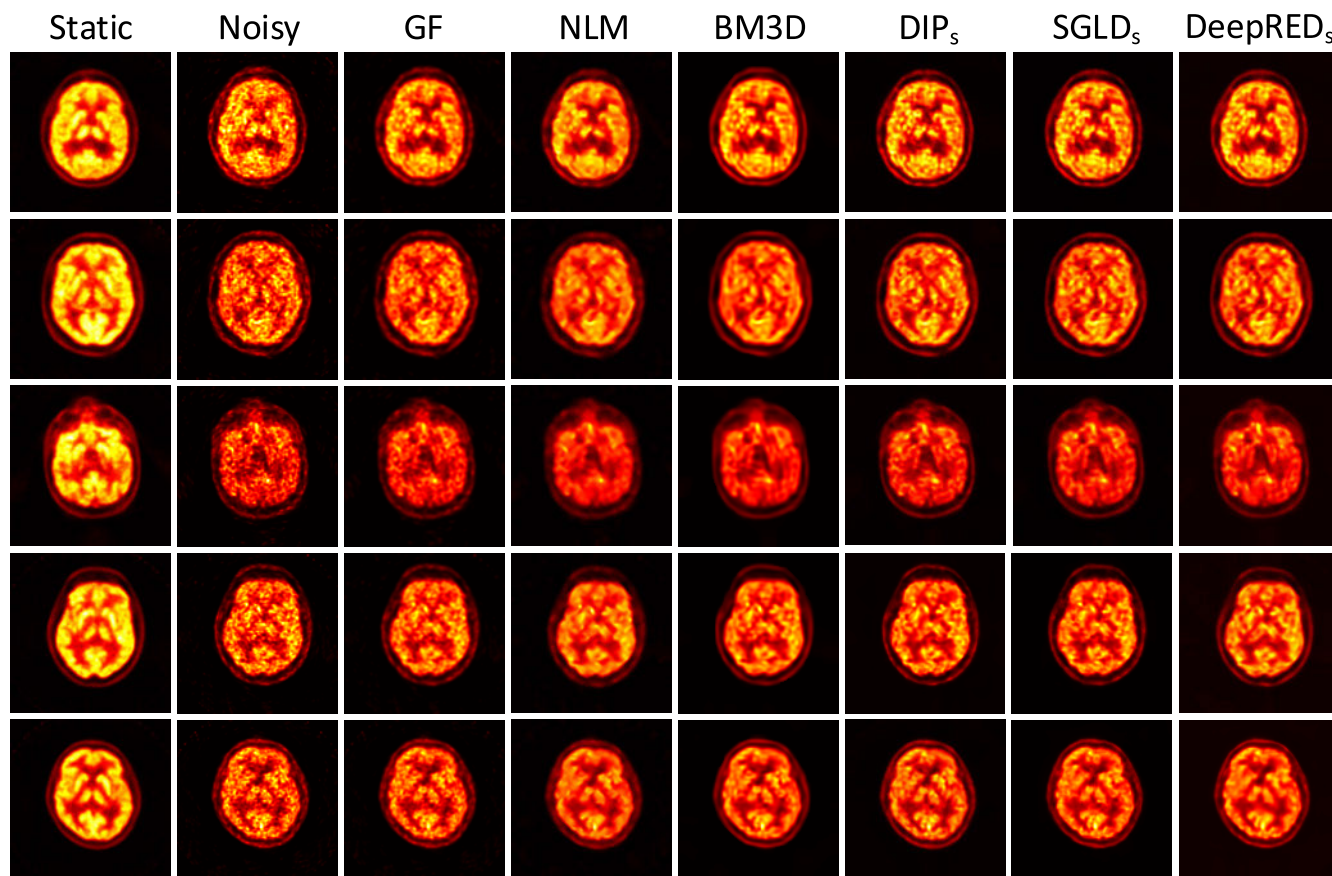


FIGURE 7. A view of approximate static PET images, original noisy images and estimated images using different methods from five patients. Each row represents the results from a patient. The first column shows approximate static PET images. The second column represents the original noisy images from different patients. Estimated images with different denoising algorithms (including GF, NLM, BM3D, DIP_s, SGLD_s and DeepRED_s) are shown from the 3rd to the 8th column. Except for the first column, the images of each row are displayed in the same grayscale range.

SGLD methods for both results from random noise network input and static image network input. Furthermore, using static image as the network input can improve the CRC of DeepRED method for gray matter region. In fact, these observations can be derived from Figure 4 by direct visual evaluation.

B. APPLICATION TO PATIENT STUDY

Following extensive validations using simulations, we applied the proposed DeepRED method to patient study (as elaborated in section IV.B). In order to provide a visual comparison of the denoised images, Figure 7 shows the denoised PET images with different methods (including GF, NLM, BM3D, DIP_s, SGLD_s, and DeepRED_s) and the corresponding approximate static PET images from five patients. Each row represents the results from a patient. The first column shows approximate static PET images. The second column shows the original images from different patients. Denoised images with different denoising algorithms (including GF, NLM, BM3D, DIP_s, SGLD_s and DeepRED_s) are shown from the 3rd to the 8th column. Since the PET scanning mechanisms of the five patients are similar, the filter parameters of the displayed images from the five patients are the same.

For GF, the FWHM was set to 3mm. For NLM, the search window was set to 2×2 . For BM3D, the standard deviation of the AWGN was set to 20. Except for the first column, the images of each row are displayed in the same grayscale range. Compared with traditional methods (GF, NLM and BM3D), DIP-based methods can reduce noise effectively while recovering more details, and the DeepRED_s method performs better than DIP_s and SGLD_s methods.

Figure 9 shows the regional CR vs. background SD curves of estimated images (generated with varying parameters). For GF, the full width at half maximum increased from 1mm to 5mm. For NLM, the search window varied from 2×2 to 6×6 . For BM3D, the standard deviation of the AWGN increased from 5 to 25. Moreover, the results from DIP-based methods were generated by varying iterations numbers. It can be seen that at the same background SD, the CR of the DeepRED method is higher than other methods, which confirms that the DeepRED method outperforms other methods.

VI. DISCUSSION

As described above, we develop a novel deep learning denoising framework to enhance the quantitative accuracy of dynamic PET image. Compared to the deep learning

TABLE 2. The Mean Value and Standard Deviation of Quantitative Evaluation Results Based on DeepRED_s with Decaying Learning Rate. The Hyperparameter $\mu = 0.2$ is Fixed and Hyperparameter λ is Changed.

	PSNR(dB)	SSIM	NRMSE
$\lambda = 0.1$	30.6792±2.76E-02	0.9010±1.21E-03	0.0292±9.29E-05
$\lambda = 0.2$	30.7134±1.39E-02	0.9026±1.17E-03	0.0291±4.66E-05
$\lambda = 0.3$	30.6762±4.61E-02	0.9027±1.34E-03	0.0293±1.56E-04
$\lambda = 0.5$	30.6263±4.59E-02	0.9003±8.59E-04	0.0294±1.56E-04
$\lambda = 0.8$	30.6023±4.26E-02	0.8995±1.05E-03	0.0295±1.44E-04

TABLE 3. The Mean Value and Standard Deviation of Quantitative Evaluation Results Based on DeepRED_s with Decaying Learning Rate. The Hyperparameter $\lambda = 0.2$ is Fixed and Hyperparameter μ is Changed.

	PSNR(dB)	SSIM	NRMSE
$\mu = 0.1$	30.6684±2.90E-02	0.9020±1.21E-03	0.0293±9.87E-05
$\mu = 0.2$	30.7134±1.39E-02	0.9026±1.17E-03	0.0291±4.66E-05
$\mu = 0.3$	30.6650±2.79E-02	0.9019±9.63E-04	0.0293±9.39E-05
$\mu = 0.5$	30.6891±2.07E-02	0.9018±1.19E-03	0.0292±6.98E-05
$\mu = 0.8$	30.6578±2.15E-02	0.8997±1.18E-03	0.0293±7.27E-05

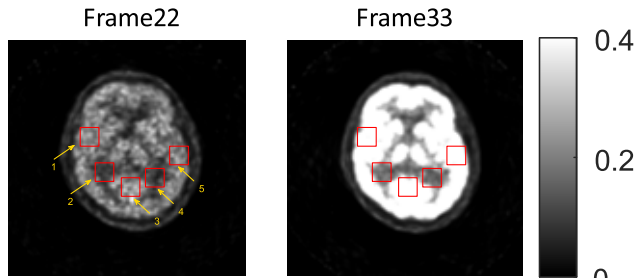


FIGURE 8. A view of the frame 22 and frame 33 from a patient. Five ROIs are drawn as red rectangles on same regions of frame 22 and frame 33.

methods based on high and low dose training pairs, the proposed method eliminates the need for tedious pre-training and potential risks of overfitting, as well as shows great performance in different data sets.

A. THE STRUCTURE OF NETWORK

We applied the skip connections to combine hierarchical features instead of concatenating for faster processing speed. Five downsampling and upsampling steps were used to learn deeper information of images. Different from [38], [41], the number of feature channels was fixed to 128 after each downsampling and upsampling step in this work. We tend to use more channels to catch more image features. Moreover, it can be found that the highest average PSNR of DIP_r with no decaying was 30.2281±1.81E-02 dB by using fixed 128 feature channel, while it was only 30.0954±2.63E-02 dB by using channels consistent with [38] and [41] through 10 tests, respectively.

B. LEARNING RATE AND OPTIMIZER

When the network reached certain iteration numbers, the output of the network may suddenly ‘collapse’ if the learning rate was too high, and it was difficult to recover the image

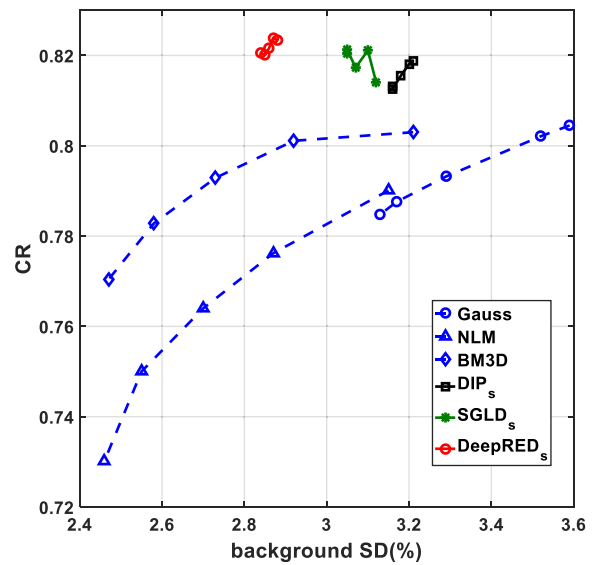


FIGURE 9. Plots of regional CR vs. background SD curves of estimated images (generated with varying parameters) using different denoising algorithms for the ROIs in Figure 8.

quality again. This phenomenon has not been described when applied to natural images. Therefore, we have tried to reduce the learning rate varying with iteration numbers. We set the initial learning rate $lr = 0.005$ decayed $lr \leftarrow 0.9 * lr$ every 1000 iterations in this work. In addition, the limited memory BFGS (L-BFGS) algorithm [57] was also implemented in our experiments. The network was optimized 100 iterations by Adam optimizer firstly and the L-BFGS optimizer was applied to the subsequent iterations. The quantitative performances of estimated images based on L-BFGS algorithm were alternating oscillation, which was difficult to choose the iteration numbers. There, the Adam optimizer was applied to all iteration.

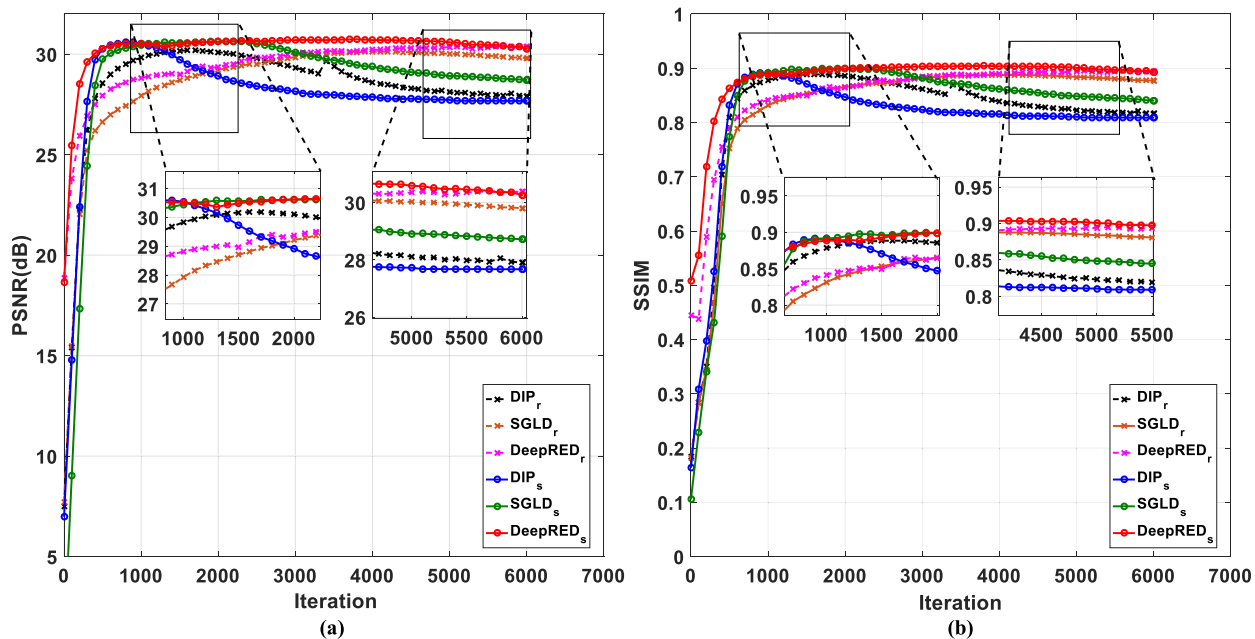


FIGURE 10. Plots of (a) PSNR and (b) SSIM curves varying with iteration number for DIP_r , $SGLD_r$, $DeepRED_r$, DIP_s , $SGLD_s$ and $DeepRED_s$ methods.

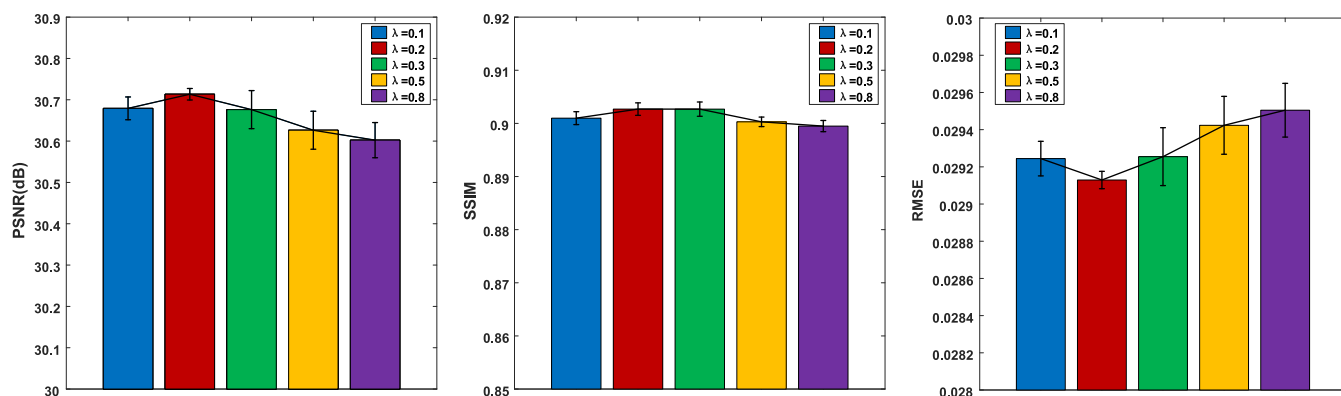


FIGURE 11. Plots of quantitative evaluation results based on $DeepRED_s$ with decaying learning rate. The hyperparameter $\mu = 0.2$ is fixed and hyperparameter λ is changed.

C. OVERFITTING

The $DeepRED$ method uses a denoising filter to regularize the DIP , which can avoid getting into local solutions and make the network learning stable. This is similar to the deep mean-shift prior (DMSP), which is used in imaging reconstruction recently [58]. Figure 10 shows the PSNR and SSIM curves varying with iteration numbers for DIP_r , $SGLD_r$, $DeepRED_r$, DIP_s , $SGLD_s$ and $DeepRED_s$ methods. As can be seen from the PSNR and SSIM curves, the DIP method is easy to overfit. This is the major drawback of the DIP method, which may be avoided by selecting the appropriate iteration number experimentally. The $SGLD$ and $DeepRED$ methods are not easy to overfit compared with DIP . Therefore, the desired image can be obtained within a wider iteration number range without worrying about the image quality degradation caused by overfitting, which provides a solution to

overcome the shortcomings of easy overfitting for DIP . [59] provided another solution idea to select the most appropriate network architecture through the neural architecture search (NAS) technique. The most appropriate iteration number may be determined by the neural search technique rather than by experience.

A point should be noted that the $SGLD$ method is slightly easier to overfit than the $DeepRED$ method in this work. As is described in section II.B, the inference procedure of the $SGLD$ method can be interpreted as a maximum likelihood estimate under a Gaussian noise model. However, the noise distribution in the PET image domain is not a standard Gaussian distribution, which is different from the inference premise of the $SGLD$ method. Besides, $SGLD$ inference includes adding noise to all parameters, including the network input z and network parameters θ . The use of

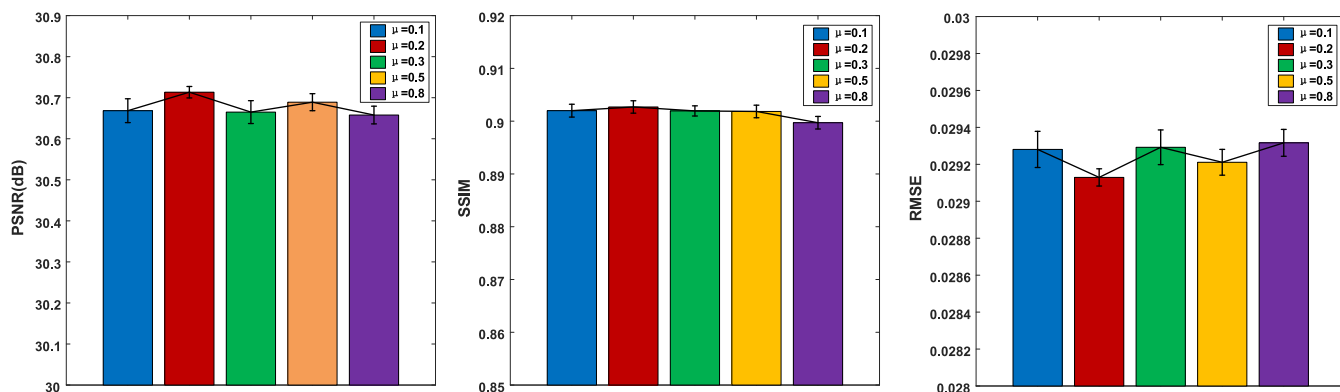


FIGURE 12. Plots of quantitative evaluation results based on DeepRED_s with decaying learning rate. The hyperparameter $\lambda = 0.2$ is fixed and hyperparameter μ is changed.

static images omits the step of adding noise to the network input, which may affect the performance of SGLD method to prevent overfitting.

D. HYPERPARAMETERS

There are two hyperparameters to be selected in the objective function (9), which have important impacts on the performance of the proposed method. In order to select the optimal parameters, the experiments based on DeepRED_s with decaying learning rates have been implemented to explore the impact of two parameters. Table 2 and Table 3 show the quantitative results of network performance when one hyperparameter is fixed and another hyperparameter is changed, and the best value of each metric is marked in black bold. For a better visual comparison, Figure 11 and Figure 12 show the results of Table 2 and Table 3 in the form of error bars, respectively. As is shown in Table 2 and Figure 12, the best quantitative results of DeepRED_s are obtained when $\lambda = 0.2$ and $\mu = 0.2$. Same results can be seen in Table 3 and Figure 12. In general, the hyperparameters $\lambda = 0.2$ and $\mu = 0.2$ were set for best performance in this paper.

VII. CONCLUSION

In this work, we develop a novel deep learning denoising framework aiming to enhance quantitative accuracy of dynamic PET images via introduction of deep image prior combined with Regularization by Denoising. The proposed method avoids the need of high quality noiseless images, and random noise or prior images can be used as the network input. Furthermore, the method combines the pre-training networks and image denoising process by constructing a constrained optimization problem and alleviates the necessity of stopping early for DIP method. Both simulation and patient data experiments show that the proposed method outperforms the conventional GF, NLM, BM3D, DIP and SGLD method in visual as well as quantitative improvements. Future work will focus on better network architecture and increasing the speed of processing in order to make the method more practical and effective in clinical data sets.

REFERENCES

- [1] A. Rahmim and H. Zaidi, "PET versus SPECT: Strengths, limitations and challenges," *Nucl. Med. Commun.*, vol. 29, no. 3, pp. 193–207, 2008.
- [2] R. M. Leahy and J. Qi, "Statistical approaches in quantitative positron emission tomography," *Statist. Comput.*, vol. 10, no. 2, pp. 147–165, 2000.
- [3] A. Rahmim, J. Tang, and H. Zaidi, "Four-dimensional (4D) image reconstruction strategies in dynamic PET: Beyond conventional independent frame reconstruction," *Med. Phys.*, vol. 36, no. 8, pp. 3654–3670, 2009.
- [4] B. Bai, Q. Li, and R. M. Leahy, "Magnetic resonance-guided positron emission tomography image reconstruction," *Seminars Nucl. Med.*, vol. 43, no. 1, pp. 30–44, 2013.
- [5] C.-M. Kao, J. T. Yap, J. Mukherjee, and M. N. Wernick, "Image reconstruction for dynamic PET based on low-order approximation and restoration of the sinogram," *IEEE Trans. Med. Imag.*, vol. 16, no. 6, pp. 738–749, Dec. 1997.
- [6] M. N. Wernick, E. J. Infusino, and M. Milosevic, "Fast spatio-temporal image reconstruction for dynamic PET," *IEEE Trans. Med. Imag.*, vol. 18, no. 3, pp. 185–195, Mar. 1999.
- [7] J. Dutta, R. M. Leahy, and Q. Li, "Non-local means denoising of dynamic PET images," *PLoS ONE*, vol. 8, no. 12, Dec. 2013, Art. no. e81390.
- [8] N. Bousson, C. C. Le Rest, M. Hatt, and D. Visvikis, "Incorporation of wavelet-based denoising in iterative deconvolution for partial volume correction in whole-body PET imaging," *Eur. J. Nucl. Med. Mol. Imag.*, vol. 36, no. 7, pp. 1064–1075, 2009.
- [9] B. T. Christian, N. T. Vandehey, J. M. Floberg, and C. A. Mistretta, "Dynamic PET denoising with HYPR processing," *J. Nucl. Med.*, vol. 51, no. 7, pp. 1147–1154, Jul. 2010.
- [10] J. Yan, J. C. Lim, and D. W. Townsend, "MRI-guided brain PET image filtering and partial volume correction," *Phys. Med. Biol.*, vol. 60, no. 3, p. 961, 2015.
- [11] Z. Bian, J. Huang, J. Ma, L. Lu, S. Niu, D. Zeng, Q. Feng, and W. Chen, "Dynamic positron emission tomography image restoration via a kinetics-induced bilateral filter," *PLoS ONE*, vol. 9, no. 2, Feb. 2014, Art. no. e89282.
- [12] K. Ote, F. Hashimoto, A. Kakimoto, T. Isobe, T. Inubushi, R. Ota, A. Tokui, A. Saito, T. Moriya, T. Omura, E. Yoshikawa, A. Teramoto, and Y. Ouchi, "Kinetics-induced block matching and 5-D transform domain filtering for dynamic PET image denoising," *IEEE Trans. Radiat. Plasma Med. Sci.*, vol. 4, no. 6, pp. 720–728, Nov. 2020.
- [13] K. Vunckx, A. Atre, K. Baete, A. Reilhac, C. M. Deroose, K. Van Laere, and J. Nuyts, "Evaluation of three MRI-based anatomical priors for quantitative PET brain imaging," *IEEE Trans. Med. Imag.*, vol. 31, no. 3, pp. 599–612, Mar. 2012.
- [14] G. Schramm, M. Holler, A. Rezaei, K. Vunckx, F. Knoll, K. Bredies, F. Boada, and J. Nuyts, "Evaluation of parallel level sets and Bowsher's method as segmentation-free anatomical priors for time-of-flight PET reconstruction," *IEEE Trans. Med. Imag.*, vol. 37, no. 2, pp. 590–603, Feb. 2018.

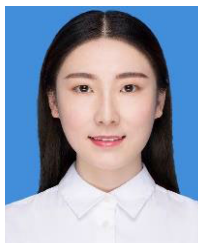
- [15] K. Baete, J. Nuyts, W. Van Paesschen, P. Suetens, and P. Dupont, "Anatomical-based FDG-PET reconstruction for the detection of hypometabolic regions in epilepsy," *IEEE Trans. Med. Imag.*, vol. 23, no. 4, pp. 510–519, Apr. 2004.
- [16] K. Baete, J. Nuyts, K. Van Laere, W. Van Paesschen, S. Ceysens, L. De Ceuninck, O. Gheysens, A. Kelles, J. Van den Eynden, P. Suetens, and P. Dupont, "Evaluation of anatomy based reconstruction for partial volume correction in brain FDG-PET," *NeuroImage*, vol. 23, no. 1, pp. 305–317, 2004.
- [17] J. Nuyts, K. Baete, D. Beque, and P. Dupont, "Comparison between MAP and postprocessed ML for image reconstruction in emission tomography when anatomical knowledge is available," *IEEE Trans. Med. Imag.*, vol. 24, no. 5, pp. 667–675, May 2005.
- [18] J. E. Bowsher, H. Yuan, L. W. Hedlund, T. G. Turkington, G. Akabani, A. Badea, W. C. Kurylo, C. T. Wheeler, G. P. Cofer, M. W. Dewhurst, and G. A. Johnson, "Utilizing MRI information to estimate F18-FDG distributions in rat flank tumors," in *Proc. IEEE Symp. Conf. Rec. Nucl. Sci.*, Oct. 2004, pp. 2488–2492.
- [19] C. Chan, R. Fulton, R. Barnett, D. D. Feng, and S. Meikle, "Postreconstruction nonlocal means filtering of whole-body PET with an anatomical prior," *IEEE Trans. Med. Imag.*, vol. 33, no. 3, pp. 636–650, Mar. 2014.
- [20] V.-G. Nguyen and S.-J. Lee, "Incorporating anatomical side information into PET reconstruction using nonlocal regularization," *IEEE Trans. Image Process.*, vol. 22, no. 10, pp. 3961–3973, Oct. 2013.
- [21] M. J. Ehrhardt, P. Markiewicz, M. Liljeroth, A. Barnes, V. Kolehmainen, J. S. Duncan, L. Pizarro, D. Atkinson, B. F. Hutton, S. Ourselin, K. Thielemans, and S. R. Arridge, "PET reconstruction with an anatomical MRI prior using parallel level sets," *IEEE Trans. Med. Imag.*, vol. 35, no. 9, pp. 2189–2199, Sep. 2016.
- [22] F. Knoll, M. Holler, T. Koesters, R. Otazo, K. Bredies, and D. K. Sodickson, "Joint MR-PET reconstruction using a multi-channel image regularizer," *IEEE Trans. Med. Imag.*, vol. 36, no. 1, pp. 1–16, Jan. 2017.
- [23] L. Lu, J. Ma, Q. Feng, W. Chen, and A. Rahmim, "Anatomy-guided brain PET imaging incorporating a joint prior model," *Phys. Med. Biol.*, vol. 60, no. 6, p. 2145, Feb. 2015.
- [24] L. Lu, N. A. Karakatsanis, J. Tang, W. Chen, and A. Rahmim, "3.5D dynamic PET image reconstruction incorporating kinetics-based clusters," *Phys. Med. Biol.*, vol. 57, no. 15, pp. 5035–5055, Aug. 2012.
- [25] K. Zhang, W. Zuo, S. Gu, and L. Zhang, "Learning deep CNN denoiser prior for image restoration," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2017, pp. 3929–3938.
- [26] O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," in *Proc. Med. Image Comput. Comput. Assist. Intervent.* Springer, 2015, pp. 234–241.
- [27] S. Ren, K. He, R. Girshick, and J. Sun, "Faster R-CNN: Towards real-time object detection with region proposal networks," in *Proc. Adv. Neural Inf. Process. Syst.*, 2015, pp. 91–99.
- [28] G. Litjens, T. Kooi, B. E. Bejnordi, A. A. A. Setio, F. Ciompi, M. Ghafoorian, J. A. W. M. van der Laak, B. van Ginneken, and C. I. Sánchez, "A survey on deep learning in medical image analysis," *Med. Image Anal.*, vol. 42, pp. 60–88, Dec. 2017.
- [29] E. Kang, J. Min, and J. C. Ye, "A deep convolutional neural network using directional wavelets for low-dose X-ray CT reconstruction," *Med. Phys.*, vol. 44, no. 10, pp. e360–e375, Oct. 2017.
- [30] S. Wang, Z. Su, L. Ying, X. Peng, S. Zhu, F. Liang, D. Feng, and D. Liang, "Accelerating magnetic resonance imaging via deep learning," in *Proc. IEEE 13th Int. Symp. Biomed. Imag. (ISBI)*, Apr. 2016, pp. 514–517.
- [31] B. Yang, L. Ying, and J. Tang, "Artificial neural network enhanced Bayesian PET image reconstruction," *IEEE Trans. Med. Imag.*, vol. 37, no. 6, pp. 1297–1309, Jun. 2018.
- [32] K. Gong, J. Guan, K. Kim, X. Zhang, J. Yang, Y. Seo, G. El Fakhri, J. Qi, and Q. Li, "Iterative PET image reconstruction using convolutional neural network representation," *IEEE Trans. Med. Imag.*, vol. 38, no. 3, pp. 675–685, Mar. 2019.
- [33] K. Gong, J. Guan, C.-C. Liu, and J. Qi, "PET image denoising using a deep neural network through fine tuning," *IEEE Trans. Radiat. Plasma Med. Sci.*, vol. 3, no. 2, pp. 153–161, Mar. 2019.
- [34] Y. Wang, B. Yu, L. Wang, C. Zu, D. S. Lalush, W. Lin, X. Wu, J. Zhou, D. Shen, and L. Zhou, "3D conditional generative adversarial networks for high-quality PET image estimation at low dose," *NeuroImage*, vol. 174, pp. 550–562, Jul. 2018.
- [35] Z. Xie, R. Baikejiang, T. Li, X. Zhang, K. Gong, M. Zhang, W. Qi, E. Asma, and J. Qi, "Generative adversarial network based regularized image reconstruction for PET," *Phys. Med. Biol.*, vol. 65, no. 12, Jun. 2020, Art. no. 125016.
- [36] A. J. Reader, G. Corda, A. Mehranian, C. da Costa-Luis, S. Ellis, and J. A. Schnabel, "Deep learning for PET image reconstruction," *IEEE Trans. Radiat. Plasma Med. Sci.*, vol. 5, no. 1, pp. 1–25, Jan. 2021.
- [37] V. Lempitsky, A. Vedaldi, and D. Ulyanov, "Deep image prior," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Jun. 2018, pp. 9446–9454.
- [38] K. Gong, C. Catana, J. Qi, and Q. Li, "PET image reconstruction using deep image prior," *IEEE Trans. Med. Imag.*, vol. 38, no. 7, pp. 1655–1665, Jul. 2019.
- [39] T. Yokota, K. Kawai, M. Sakata, Y. Kimura, and H. Hontani, "Dynamic PET image reconstruction using nonnegative matrix factorization incorporated with deep image prior," in *Proc. IEEE/CVF Int. Conf. Comput. Vis. (ICCV)*, Oct. 2019, pp. 3126–3135.
- [40] J. Cui, K. Gong, N. Guo, C. Wu, X. Meng, K. Kim, K. Zheng, Z. Wu, L. Fu, B. Xu, Z. Zhu, J. Tian, H. Liu, and Q. Li, "PET image denoising using unsupervised deep learning," *Eur. J. Nucl. Med. Mol. Imag.*, vol. 46, no. 13, pp. 2780–2789, Dec. 2019.
- [41] F. Hashimoto, H. Ohba, K. Ote, A. Teramoto, and H. Tsukada, "Dynamic PET image denoising using deep convolutional neural networks without prior training datasets," *IEEE Access*, vol. 7, pp. 96594–96603, 2019.
- [42] F. Hashimoto, H. Ohba, K. Ote, A. Kakimoto, H. Tsukada, and Y. Ouchi, "4D deep image prior: Dynamic PET image denoising using an unsupervised four-dimensional branch convolutional neural network," *Phys. Med. Biol.*, vol. 66, no. 1, Jan. 2021, Art. no. 015006.
- [43] Z. Cheng, M. Gadelha, S. Maji, and D. Sheldon, "A Bayesian perspective on the deep image prior," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2019, pp. 5443–5451.
- [44] M. Welling and Y. W. Teh, "Bayesian learning via stochastic gradient Langevin dynamics," in *Proc. ICML*, 2011, pp. 681–688.
- [45] G. Mataev, M. Elad, and P. Milanfar, "DeepRED: Deep image prior powered by RED," 2019, *arXiv:1903.10176*. [Online]. Available: <http://arxiv.org/abs/1903.10176>
- [46] G.-H. Chen, J. Tang, and S. Leng, "Prior image constrained compressed sensing (PICCS): A method to accurately reconstruct dynamic CT images from highly undersampled projection data sets," *Med. Phys.*, vol. 35, no. 2, pp. 660–663, Jan. 2008.
- [47] F. Hashimoto, H. Ohba, K. Ote, and H. Tsukada, "Denoising of dynamic sinogram by image guided filtering for positron emission tomography," *IEEE Trans. Radiat. Plasma Med. Sci.*, vol. 2, no. 6, pp. 541–548, Nov. 2018.
- [48] S. Boyd, "Distributed optimization and statistical learning via the alternating direction method of multipliers," *Found. Trends Mach. Learn.*, vol. 3, no. 1, pp. 1–122, 2010.
- [49] R. M. Neal, *Bayesian Learning for Neural Networks*. New York, NY, USA: Springer-Verlag, 1996.
- [50] Y. Romano, M. Elad, and P. Milanfar, "The little engine that could: Regularization by denoising (RED)," *SIAM J. Imag. Sci.*, vol. 10, no. 4, pp. 1804–1844, Jan. 2017.
- [51] M. V. Afonso, J. M. Bioucas-Dias, and M. A. T. Figueiredo, "Fast image recovery using variable splitting and constrained optimization," *IEEE Trans. Image Process.*, vol. 19, no. 9, pp. 2345–2356, Sep. 2010.
- [52] C. Godard, O. M. Aodha, and G. J. Brostow, "Unsupervised monocular depth estimation with left-right consistency," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2017, pp. 270–279.
- [53] F. Varela, J.-P. Lachaux, E. Rodriguez, and J. Martinerie, "The brainweb: Phase synchronization and large-scale integration," *Nature Rev. Neurosci.*, vol. 2, no. 4, pp. 229–239, Apr. 2001.
- [54] G. Wang and J. Qi, "PET image reconstruction using kernel method," *IEEE Trans. Med. Imag.*, vol. 34, no. 1, pp. 61–71, Jan. 2015.
- [55] K. Dabov, A. Foi, V. Katkovnik, and K. Egiazarian, "Image denoising with block-matching and 3D filtering," *Proc. SPIE*, vol. 6064, Feb. 2006, Art. no. 606414.
- [56] C. R. Jack *et al.*, "The Alzheimer's disease neuroimaging initiative (ADNI): MRI methods," *J. Magn. Reson. Imag., Off. J. Int. Soc. Magn. Reson. Med.*, vol. 27, no. 4, pp. 685–691, 2008.
- [57] C. Zhu, R. H. Byrd, P. Lu, and J. Nocedal, "Algorithm 778: L-BFGS-B: Fortran subroutines for large-scale bound-constrained optimization," *ACM Trans. Math. Softw.*, vol. 23, no. 4, pp. 550–560, 1997.

- [58] M. Zhang, M. Li, J. Zhou, Y. Zhu, S. Wang, D. Liang, Y. Chen, and Q. Liu, "High-dimensional embedding network derived prior for compressive sensing MRI reconstruction," *Med. Image Anal.*, vol. 64, Aug. 2020, Art. no. 101717.
- [59] K. Ho, A. Gilbert, H. Jin, and J. Collomosse, "Neural architecture search for deep image prior," 2020, *arXiv:2001.04776*. [Online]. Available: <http://arxiv.org/abs/2001.04776>



HAO SUN received the B.Eng. degree from Southern Medical University, in 2019, where he is currently pursuing the master's degree in engineering with the School of Biomedical Engineering. His research interests include artificial intelligence and PET imaging.

LIHONG PENG received the B.Sc. degree in automation from the Hubei University of Technology, Wuhan, China, in 2007. She is currently pursuing the Ph.D. degree with the School of Biomedical Engineering, Southern Medical University, Guangzhou, China. Her research interest includes medical image reconstruction and analysis.



HONGYAN ZHANG received the B.Eng. degree from Southern Medical University, in 2018, where she is currently pursuing the master's degree with the Guangdong Provincial Key Laboratory of Medical Image Processing, School of Biomedical Engineering. Her research interest includes dynamic PET factor analysis.



YURU HE received the B.Eng. degree from Southern Medical University, in 2018, where she is currently pursuing the master's degree in engineering with the Guangdong Provincial Key Laboratory of Medical Image Processing, School of Biomedical Engineering. Her research interest includes PET/MRI denoising.



SHUANGLIANG CAO received the B.Eng. degree from the Hebei University of Science and Technology, in 2012. He is currently pursuing the Ph.D. degree with the Guangdong Provincial Key Laboratory of Medical Image Processing, School of Biomedical Engineering, Southern Medical University. His research interests include dynamic PET imaging and parametric imaging.



LIJUN LU received the Ph.D. degree in biomedical engineering from Southern Medical University, in 2012. He is currently a Full Professor with the Guangdong Provincial Key Laboratory of Medical Image Processing, School of Biomedical Engineering, Southern Medical University. His research interests include PET imaging methods, medical image processing, and radiomics.

...