IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received February 6, 2021, accepted February 11, 2021, date of publication March 9, 2021, date of current version March 22, 2021.

Digital Object Identifier 10.1109/ACCESS.2021.3064926

Dynamic PET Image Denoising With Deep Learning-Based Joint Filtering

YURU HE[1,2], SHUANGLIANG CAO[1,2], HONGYAN ZHANG[1,2], HAO SUN[1,2], FANGHU WANG[3], HUOBIAO ZHU[1,2], WENBING LV[1,2], AND LIJUN LU[1,2]

[1]School of Biomedical Engineering, Southern Medical University, Guangzhou 510515, China
[2]Guangdong Provincial Key Laboratory of Medical Image Processing, Southern Medical University, Guangzhou 510515, China
[3]WeiLun PET Center, Department of Nuclear Medicine, Guangdong Provincial People's Hospital and Guangdong Academy of Medical Sciences, Guangzhou 510080, 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.

ABSTRACT Dynamic positron emission tomography (PET) imaging usually suffers from high statistical noise due to low counts of the short frames. This study aims to improve the image quality of the short frames by utilizing information from other modality. We develop a deep learning-based joint filtering framework for simultaneously incorporating information from longer acquisition PET frames and high-resolution magnetic resonance (MR) images into the short frames. The network inputs are noisy PET images and corresponding MR images while the outputs are linear coefficients of spatially variant linear representation model. The composite of all dynamic frames is used as training label in each sample, and it is down-sampled to 1/10th of counts as the training input. L1-norm combined with two gradient-based regularizations constitute the loss function during training. Ten realistic dynamic PET/MR phantoms based on BrainWeb are used for pre-training and eleven clinical subjects from Alzheimer's Disease Neuroimaging Initiative further for fine-tuning. Simulation results show that the proposed method can reduce the statistical noise while preserving image details and achieve quantitative enhancements compared with Gaussian, guided filter, and convolutional neural network trained with the mean squared error. The clinical results perform better than others in terms of the mean activity and standard deviation. All of the results indicate that the proposed deep learning-based joint filtering framework is of great potential for dynamic PET image denoising.

INDEX TERMS Positron emission tomography, convolution neural network, denoising, spatially variant linear representation model, joint filtering.

I. INTRODUCTION

Positron emission tomography (PET) is an *in vivo* functional imaging modality reflecting the metabolism in a living body by injecting target-specific PET tracers. Though the physical characteristics are continuing to develop, dynamic PET imaging still suffers from high statistic noise due to low counts of the short frames [1]. To improve the quality of PET images, high-dose tracer injection to tissue is generally administered in clinical practice, which increases the risk of radiation exposure to patients and technicians undoubtedly. Therefore, it's necessary to improve the quality of low-count PET images.

The associate editor coordinating the review of this manuscript and approving it for publication was Larbi Boubchir.

Independent-frame 3D image reconstruction is commonly accomplished by using statistical image reconstruction methods, such as maximum likelihood expectation-maximization (MLEM) methods. However, MLEM exhibits high variance with increasing iteration at low counts [2]. This low-count is further accentuated with increased temporal sampling. Reconstruction based method commonly take this ill-posedness inherent in PET using Bayesian methods through the introduction of prior model [3]–[5]. On the other hand, post-reconstruction methods also attracted more attention considering their direct application to the clinic [6]–[8].

Among the post-reconstruction method, the Gaussian filter is most commonly applied for PET image filtering, which reduces the noise with edge blurring. The edge-preserving bilateral filter (BF) [9] allows both increasing

the signal-to-noise ratio (SNR) of clinical PET image and preventing smoothing-induced underestimation of maximum standardized uptake value in small lesions [10]. However, the BF may suffer from ''gradient reversal'' artifacts, which finally results in the unwanted profiles exhibit around edges. The guided filter [11] proposed by He *et al*. can be used as an edge-preserving smoothing operator and avoids the ''gradient reversal'' artifacts. For the guided filter, the guided images are very critical during the filter process. Lu *et al*. proposed the composite image guided filter technology for dynamic PET images denoising [12], where the composite image was the sum of dynamic frames. F. Hashimoto *et al*. presented a sinogram-based dynamic image guided filtering (SDIGF) algorithm for PET denoising, where the guided image was the normalized static PET sinogram [13]. Furthermore, anatomical information derived from magnetic resonance (MR) images or computed tomography (CT) images have also been incorporated as the guided image. Yan *et al.* proposed an MR-guided brain PET image filter and then incorporated partial volume effects into the model, which showed good results in terms of visual inspection and quantitative metrics [14]. However, the methods incorporated the anatomical images may introduce irrelevant information into PET images due to the mismatch of the structures. Recently, Pan *et al*. [15] proposed a new joint filter based on the spatially variant linear representation model (SVLRM) which is an improvement of the guided filter. It shows that the joint filter can transfer the meaningful structural details of the guided images and input images to the target image. The detailed description is presented in Section II (A).

Deep neural networks (DNNs) have been widely and successfully used in computer vision tasks, such as object tracking [16], image segmentation [17], and image classification [18]. Inspired by this, many studies have investigated PET images denoising by using DNNs, superior performance and faster speed have been achieved compared to conventional methods [19]–[21]. However, there are three following problems in denoising PET images with DNNs.

- Lack of high-quality PET images as the training label. In the clinic, high-quality PET images are usually obtained by high-dose tracer injection which is harmful to human health. Cui *et al*. proposed an unsupervised deep learning method using the noisy PET image itself as the training label and CT/MR images as input [19].
- Lack of enough clinical data to train the network well. The networks are likely to overfit the training datasets and obtain poor testing results when using insufficient training data. Hashimoto *et al*. proposed the PET image denoising method based on the deep image prior, which only used a single data pair for training [21].
- Mean squared error (MSE) based loss function which is commonly used usually results in blurry outputs. Kaplan *et al*. combined the gradient and total variation as the specific characteristics with MSE to reduce blurry [22]. Gong *et al.* used the perceptual loss which comparing image feature maps extracted from the pre-trained network instead of pixel intensities [23].

In this work, we proposed a deep learning-based joint filtering framework to improve the image quality of low-count dynamic PET scanning. The main contributions of this work are as follows.

- The high-quality training label images were derived from the composited of all dynamic frames, and down-sampled to obtain the network training inputs. This method of data acquisition only needs a single dynamic scanning of each patient.
- The network was pre-trained with simulated data and then the last two convolution layers were fine-tuned with clinical data. A similar idea was presented in [23], where they used clinical data to fine-tune the last two convolution layers and the last residual block.
- The loss function was combined the L1-norm with the edge-preserving and structure-preserving features. Through minimizing the Manhattan distance and the gradient difference between labels and outputs, and maximizing the gradient of the result images, noise reduction and structural details preservation were achieved.

II. METHODS

A. GUIDED FILTER AND SVLRM

The key assumption of guided filter is a local linear model between the output image Q_i and the guided image G_i in a window ω_k centered at the voxel k:

$$Q_i = a_k G_i + b_k, \quad \forall i \in \omega_k, \tag{1}$$

where a_k and b_k are linear coefficients that mapping G_i to Q_i in ω_k. This local linear model ensures that the structures of the guidance image G_i are directly transferred to the output image Q_i. Image noise n_i is defined as the difference between the output image Q_i and the input image I_i:

$$n_i = Q_i - I_i. \tag{2}$$

In order to minimize the noise at voxel i, the cost function E can be obtained while satisfying the constraint of (1) in the local window ω_k:

$$E(a_k, b_k) = \sum_{i \in \omega_k} ((a_k G_i + b_k - I_i)^2 + \varepsilon a_k^2), \tag{3}$$

where a_k and b_k assume to be constant in the local window ω_k, ε is the regularization coefficient to penalize constraint a_k^2. The voxel i in the output image is obtained by averaging the overlapping windows in which included the voxel i and is given by:

$$Q_i = \frac{1}{|\omega|} \sum_{k \in \omega_i} (a_k G_i + b_k) = \overline{a}_i G_i + \overline{b}_i, \tag{4}$$

where $|\omega|$ is the number of windows including voxel i, $\overline{a}_i = \frac{1}{|\omega|} \sum_{k \in \omega_i} a_k$ and $\overline{b}_i = \frac{1}{|\omega|} \sum_{k \in \omega_i} b_k$ are the average coefficients of all the local windows including the voxel i.

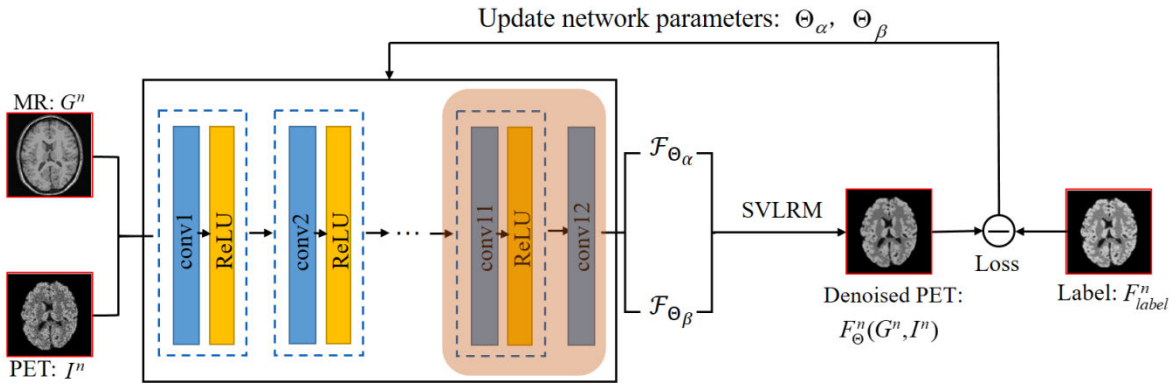


FIGURE 1. Schematic of the proposed deep learning-based joint filtering framework.

Although the guided filter has been proved effective in many applications. There exist two problems as follows.

- Due to the average operation in (4), high-frequency information in the local image patch will be suppressed.
- $\nabla Q_i \approx \overline{a}_i \nabla G_i$ makes the structures of guided image transfer into the output image, which may introduce extraneous details.

To address the problem, Pan *et al.* [15] proposed a new joint filter based on the spatially variant linear representation model (SVLRM):

$$Q_i = \alpha(G_i, I_i)G + \beta(G_i, I_i), \tag{5}$$

where $\alpha(G_i, I_i)$ and $\beta(G_i, I_i)$ are the spatially variant linear representation coefficients, which are determined by the guided image G and input image I. Then a convolutional neural network (CNN) was developed to estimate the spatially variant linear representation coefficients. Inspired by this, we proposed a deep learning-based joint filtering framework to reduce the noise in the simulated and real PET images while preserving edges and introducing the meaningful structure from the MR images and PET noisy images to the target images.

B. NETWORK ARCHITECTURE

Based on the SVLRM, we used the CNN with 12 convolution layers to denoise the PET noisy image. The schematic diagram of the CNN architecture is shown in Fig. 1. The inputs of CNN are PET noisy image I^n and corresponding MR image G^n, and the label is high-count PET image F^n_{label}. $\{I^n, G^n, F^n_{label}\}_{n=1}^N$ indicates a set of N training samples. The outputs of CNN are the linear coefficients $\mathcal{F}_{\Theta_\alpha}(G^n, I^n)$ and $\mathcal{F}_{\Theta_\beta}(G^n, I^n)$ of the SVLRM, where Θ_α and Θ_β are network parameters respectively. The denoised PET image is defined as:

$$\mathcal{F}_\Theta(G^n, I^n) = \mathcal{F}_{\Theta_\alpha}(G^n, I^n)G^n + \mathcal{F}_{\Theta_\beta}(G^n, I^n). \tag{6}$$

The loss between the denoised PET image and the training label was calculated, and then backpropagation of loss to the network, using optimization algorithms to update network parameters. The size of the convolution kernel in each layer is 3×3 pixels, and the stride value is 1. To ensure that the image size is unchanged after network processing, we added one zero-padding to the inputs of each layer. In this situation, our network can process images of any size. The number of features at the first 11 convolution layers is set to be 64. The input dimension of the network is $64 \times 64 \times 2 \times 10$, where the patch size is 64×64 and the batch size is 10. Flip and rotate randomization was used in image patches to amplify data. The activation function ReLU was used following each convolution layer except for the final convolution layer. Pooling and batch normalization (BN) layers were not included.

C. LOSS FUNCTION

As a frequently used loss function, MSE quantifying the error between the training label F^n_{label} and network output $\mathcal{F}_\Theta$, but it has been reported that MSE usually results in blurry outputs. Therefore, we proposed to use L1-norm and gradient-based regularizations as the loss function in the training process, which is defined as:

$$\begin{aligned} &L(\mathcal{F}_\Theta, F^n_{label}) \\ &= \frac{1}{N}\sum_{n=1}^{N} \left\| \mathcal{F}_\Theta - F^n_{label} \right\|_1 - w_1\left(\frac{1}{N}\sum_{n=1}^{N} \|\nabla \mathcal{F}_\Theta\|_2^2\right) \\ &\quad + w_2\left(\frac{1}{N}\sum_{n=1}^{N} \left\|\nabla \mathcal{F}_\Theta - \nabla F^n_{label}\right\|_2^2\right). \end{aligned} \tag{7}$$

The first term is the L1-norm, which is minimized to ensure the label image and network output are similar in values. Since the L1-norm is nondifferentiable, the Charbonnier penalty function $\rho(x) = \sqrt{x^2 + \delta^2}$ was used to approximate it. The second term is the edge-preserving feature, based on the total variation of network outputs, which is maximized in loss function to preserve the structure details The third term is the structure-preserving feature, which is minimized to induce the structural components of the network output patches are as similar as possible to those of the label image patches. The loss function aims to reduce the statistic noise

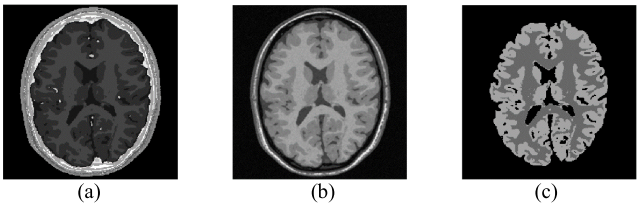


FIGURE 2. Transaxial slices based on BrainWeb used for simulation from the same subject. (a) the discrete MR image from slice 34; (b) the T1-weight MR image from slice 34; (c) The ground-truth PET image from the 24th frame of slice 34 which was simulated from the corresponding discrete MR image as Fig. 2(a).

while preserving structural details and edges in the network output.

D. FINE-TUNING

CNN requires large training datasets to work well. Due to the limited number of clinical datasets, we combined the simulated brain data with real data to train the network in this study. As the structure of the simulated data is similar to those of the clinical data, we considered that the low-level features trained from simulated data are similar to those from clinical data. During the training, we pre-trained the entire network with simulated datasets, followed by fine-tuning the last two convolution layers only with clinical data. The fine-tune area is shown in the pink shadow region in Fig. 1.

III. EXPERIMENT DESIGN

A. SIMULATION STUDY

1) MR SIMULATION

For MR simulation, we simulated T1-weighted MR images and the discrete images based on the BrainWeb database [24]. Each discrete image consists of 11 individual regions of the brain (gray matter, white matter, fat, muscle/skin, skull, vessels, connective, dura matter, bone marrow, cerebrospinal fluid and background). The matrix dimensions of the discrete MR images were 362 × 434 × 362 and that of the T1-weighted MR images were 256 × 256 × 181 with 1-mm isotropic voxel size [25]. Through image cropping, linear transformation, and partial slices extraction, the dimensions of discrete MR images were equal to the T1-weighted MR images while the structures of these were roughly matched, as shown in Fig. 2 (a) and (b).

2) PET SIMULATION

PET images were obtained by extracting gray matter and white matter from the MR discrete brain images and assigning the activity values. The regional time-activity curves of gray matter and white matter, based on the glucose metabolism of ^{18}F-fluorodeoxyglucose (^{18}F-FDG), were consistent with those used in [26], as shown in Fig. 3. The scanning schedule including 24-time frames for 60 min: 4 × 20 s, 4 × 40 s, 4 × 60 s, 4 × 180 s, and 8 × 300 s. The 24th frame of the PET image is shown in Fig. 2 (c). We performed realistic analytic simulation for the geometry of the GE Discovery ST PET/CT scanner of which the system matrix

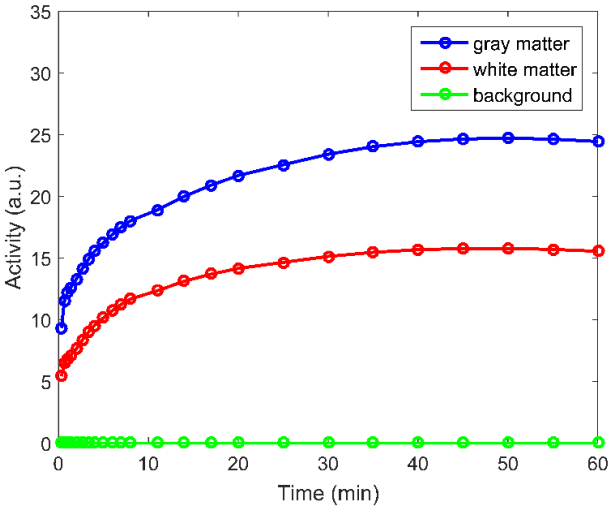


FIGURE 3. Regional time-activity curves of gray matter and white matter used in the simulation studies.

was shaped by the parallel strip line integration method. The matrix dimensions of the PET images were 256 × 256 × 181 and the cubic voxel sizes of 1.25 × 1.25 × 1.25 mm^3. Noise-free projection data which consisted of 256 bins, 192 angles, and 181 slicers were generated by forward projecting the dynamic PET images. To simulate the partial volume effect of the PET images, we used the Gaussian filter with full-width-half-maximum (FWHM) equal to 3.5 pixels to blur the noise-free sinogram. The expected total number of events over 60 min was 100 M. We added uniform random events that value equal to 20% of simulated counts to simulate background events. The attenuation and object-dependent scatter were not included. Finally, we generated a set of 10 noisy realizations by introducing random Poisson noise to the sinogram. PET images were reconstructed using the MLEM with 120 iterations. The simulated method of system matrix and reconstruction were referred the image reconstruction toolbox (IRT) which proposed by Fessler and his group [27].

3) TRAINING AND TESTING DATA

Ten 3D brain phantoms were used in the simulation, where nine phantoms were used for training and one for testing. After discarding axial slices at the two ends with low activity, the matrix dimensions of PET and MR images were 256 × 256 × 75. To reduce the computational costs and increase the data for training, we extracted the patches from the training input and label at the same positions with 64 × 64 pixels randomly. A total of 6750 training pairs were used for network training, including 9 (number of phantoms) × 75 (number of axial slices extracted from each phantom) × 10 (number of patches extracted from each slice). To obtained the training label, we summed the entire dynamic PET data into one static frame and reconstructed it. The corresponding noisy input was obtained by downsampling the label data randomly to 1/10th of counts and reconstructing. Fig. 4 shows three training input and label image pairs from different brain

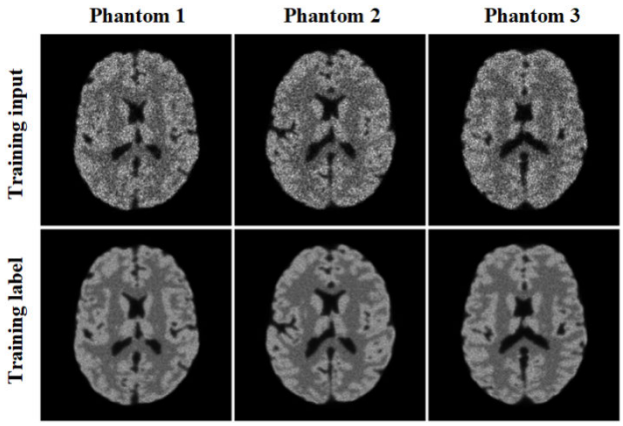


FIGURE 4. Three training image pairs from slice 34 of different simulated brain phantoms. Each column represents the same phantom. The images at the top row are the training inputs while at the bottom row are the corresponding labels.

phantoms. For testing, the last frame of dynamic PET images and the corresponding MR images both with 256 × 256 pixels were used as the network inputs in the simulation study.

4) FIGURES OF MERIT

To compare the images enhanced from the different algorithms described in the previous section, we used quantitative evaluation criteria involving ensemble variance and means squared bias. The ensemble variance was defined as

$$\text{Vars} = \frac{1}{R}\frac{\sum_{r=1}^{R}\sum_{j=1}^{N}(x_j^r - \overline{x}_j)^2}{\sum_{j=1}^{N}(x_j^{\text{true}})^2}, \tag{8}$$

where N is the total number of voxels, R is the number of noise realizations. $\overline{x}_j = \frac{1}{R}\sum_{r=1}^{R} x_j^r$ represents the ensemble mean of the denoised images at voxel j, x_j^{true} denotes the ground truth PET image value at voxel j, x_j^r denotes the denoised PET image value at voxel j in the rth noise realization. The ensemble mean squared bias was defined as

$$\text{Bias}^2 = \frac{\sum_{j=1}^{N}(\overline{x}_j - x_j^{\text{true}})^2}{\sum_{j=1}^{N}(x_j^{\text{true}})^2}, \tag{9}$$

where the variables were defined as those in (8). We also computed the contrast recovery coefficient (CRC) of the region of interest (ROI) versus standard deviation (STD) in the background. The CRC was defined as

$$\text{CRC} = \frac{1}{R}\sum_{r=1}^{R}\left(\frac{\bar{x}_a^r}{\bar{x}_b^r} - 1\right)\Big/\left(\frac{\bar{x}_a^{\text{true}}}{\bar{x}_b^{\text{true}}} - 1\right), \tag{10}$$

where $\bar{x}_a^r$ denotes the mean activity from ROI in realization r and $\bar{x}_b^r$ denotes the mean activity from the background in realization r. $\bar{x}_a^{\text{true}}$ and $\bar{x}_b^{\text{true}}$ denotes the mean activity from ROI and background of the ground truth PET image respectively. The background STD was defined as

$$\text{STD} = \frac{1}{N_b}\sum_{j=1}^{N_b}\frac{\sqrt{\frac{1}{R-1}\sum_{r=1}^{R}(x_b^{r,j} - \overline{x}_b^j)^2}}{\overline{x}_b^j}. \tag{11}$$

Here $x_b^{r,j}$ denotes the activity value of the background in denoised PET image at voxel j in the rth noise realization. $\overline{x}_b^j$ represents the mean activity of background at voxel j over 10 realizations, and N_b is the number of voxels from the background. The gray matter was the ROI and the white matter was chosen as the background in this work. Furthermore, the root means squared error (RMSE) and structural similarity (SSIM) also be used as evaluation metrics.

B. PATIENT STUDY

After pre-training the network using digital phantoms, we fine-tuned the last two convolution layers using real brain data. The dataset contains 11 subjects, each subject including PET and T1-weighted MR images. Ten subjects were used for the fine-tuning, and one subject was reserved for testing. The cross-validation was performed that one of eleven subjects take a turn as the test dataset and the rest ten for training. All data were obtained from the healthy normal controls of the Alzheimer's Disease Neuroimaging Initiative (ADNI) database [28]. The MR images were acquired with T1-weighted MP-RAGE sequence. The matrix size was 256 × 256 × 170, and the voxel size was 0.94 × 0.94 × 1.2 mm^3. Dynamic PET scanning was performed for 60 min scan after the injection of ^{18}F-FDG at a dose of 5 ± 0.5 mCi. The dynamic PET data consisted of 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. The PET matrix size was 128 × 128 × 63, and the voxel size was 2.1 × 2.1 × 2.4 mm^3. Each MR image was registered with the corresponding PET image using 3D slicer [29]. After discarding axial slicers at the two ends with low activity, the PET and MR image array size was 128 × 128 × 36. Due to the clinical sinogram was not available, we roughly simulated the system matrix and further calculated the sinogram with Fessler's IRT toolbox according to the data description from ADNI.

In the fine-tuning procedure, images reconstructed using the entire dynamic PET data were treated as the training labels, and images reconstructed using 1/10th of counts were treated as noisy input. For clinical testing, the last frame of dynamic PET images and the corresponding MR images both with 128 × 128 pixels were used as the network inputs. Like the simulated experiment, the image patches with 64 × 64 pixels were randomly extracted from the PET image and the MR image at the same positions for fine-tuning. A total of 3600 training pairs were used for network fine-tune, including 10 (number of subjects) × 36 (number of axial slices extracted from each subject) × 10 (number of patches extracted from each slice).

To evaluate the performance of the proposed method, we randomly selected twenty square ROIs with the size of 12 × 12 mm^2 in each denoised image. Since the true activity values of the clinical image are unknown, the bias cannot be calculated. Instead, the mean activity of ROIs and the mean STD of these ROIs were calculated, where the STD of each ROI was similar to (11).

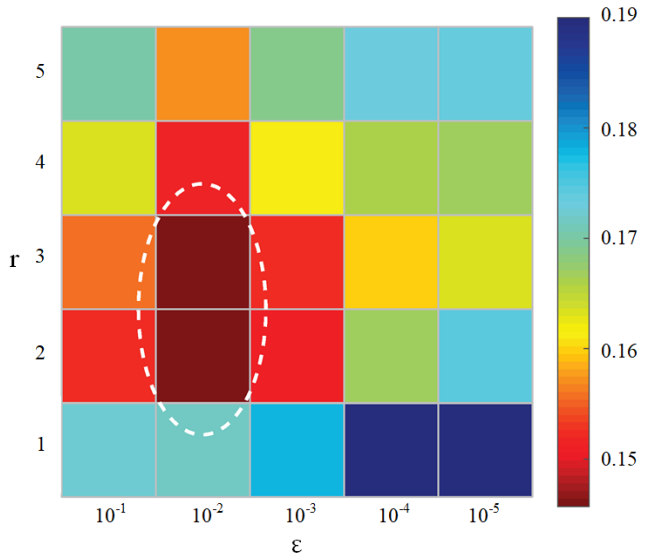


FIGURE 5. Heatmap of the mean value of RMSE for all test data processed by the image guided filter.

IV. RESULTS

We evaluated the proposed method in comparison with post-filtering with Gaussian filter and guided filter, and the CNN trained with MSE as loss function (CNN-MSE). The overall performance of all the test data was evaluated and then two slices were selected for visual comparison. For one slice, the ensemble variance versus ensemble means squared bias and CRC vs. STD curves over the 10 noise realizations were compared by varying the number of reconstruction iterations and parameters of methods, respectively.

A. SIMULATION STUDY

1) PARAMETER OPTIMIZATION

The performance of filter methods (including Gaussian filter, guided filter, CNN-MSE, proposed method) strongly depends on the parameters. First, the radius r of the local window and the regularization parameter ε in the guided filter were optimized using RMSE for all slices as shown in Fig. 5. The radius r ranged from 1 to 5, and the regularization parameter ε ranged from 10^{-1} to 10^{-5}. Fig. 5 shows that the optimized parameters can be found in the dark red region (with the minimum RMSE), where the value of r is ranged from 2 to 3, and the value of ε is 10^{-2}. The radius r and the regularization parameter ε were set to 2 and 10^{-2}. The FWHM of the Gaussian filter was set to 2.5 voxels empirically, and the number of training iteration was set to 200 epochs in both CNN-MSE and the proposed method. All images were reconstructed using MLEM with 120 iterations.

The proposed convolutional neural network was implemented using Caffe, which is a deep learning platform made with expression, modularity, and speed. The training optimizer was the Adam algorithm with parameters $\beta_1 = 0.9$, $\beta_2 = 0.999$, and $\eta = 10^{-4}$. The learning rate was initialized as 10^{-7} and trained 200 epochs with simulated data. For fine-tuning, the learning rate was initialized as 10^{-6} and trained 50 epochs with clinical data. The batch size was set to be 10 in all training.

TABLE 1. Quantitative evaluations on the all simulated test dataset in terms of the average RMSEs and average SSIMs.

	Avg. RMSEs	Avg. SSIMs
Input	0.322 ± 0.040	0.510 ± 0.070
Gaussian	0.185 ± 0.024	0.677 ± 0.060
Guided Filter	0.149 ± 0.014	0.741 ± 0.041
CNN-MSE	0.150 ± 0.016	0.774 ± 0.037
Proposed	0.138 ± 0.016	0.805 ± 0.034

Values are expressed as means ± standard deviations.

The selection of proper parameters (w_1, w_2, δ) can be critical for the cost function. In this work, we utilized a simple yet practical approach named the control variable. We set the value range of each parameter empirically and optimized one parameter with average RMSEs and average SSIMs on the simulated test data while others were fixed [22]. For example, the w_1 was ranged from 0.000005 to 0.05 and w_2 was 0.075 and δ was 10^{-6}. When w_1 was 0.0005, the proposed method has the lowest average RMSE and the highest average SSIM, as shown in Fig. 6. Thus, 0.0005 was chosen as the w_1 value. A similar strategy was adopted in the w_2 and δ optimization. The optimal w_2 value was 0.075 when w_1 was 0.0005 and δ was 10^{-6}, as shown in Fig. 7. In Fig. 8, the proposed method can achieve the same lowest RMSE value when δ was 10^{-6} and 10^{-5}, but the highest SSIM was achieved by 10^{-6}. Thus, we chose 10^{-6} as the optimal δ. However, no matter the parameter changed, the proposed method can keep the lowest average RMSE and the highest average SSIM compared with the noisy input, Gaussian, guided filter, and the CNN-MSE, as shown in Fig. 6-8.

2) PERFORMANCE COMPARISON

To provide the overall performance evaluation, as shown in TABLE 1, we calculated the average RMSEs and average SSIMs of the test dataset, it can be seen that the proposed method obtained the lowest RMSE and highest SSIM where the RMSE was reduced by 56.8% and the SSIM was increased by 57.5% compared with the input images. Fig. 9 shows the quantitative evaluation of each test image, and it can be seen that the proposed method also has favorable performance against the compared methods in each testing image. To provide a visual evaluation, Fig. 10 shows the transverse view of slice 1 and slice 34 from the test phantom and image processed by different filter methods. The proposed method can simultaneously preserve the edge and reduce statistical noise compared to other methods. This can be clearly seen in the rectangular red box. The guided filter introduces the extraneous structures, which can be seen at the red arrow in Fig. 10.

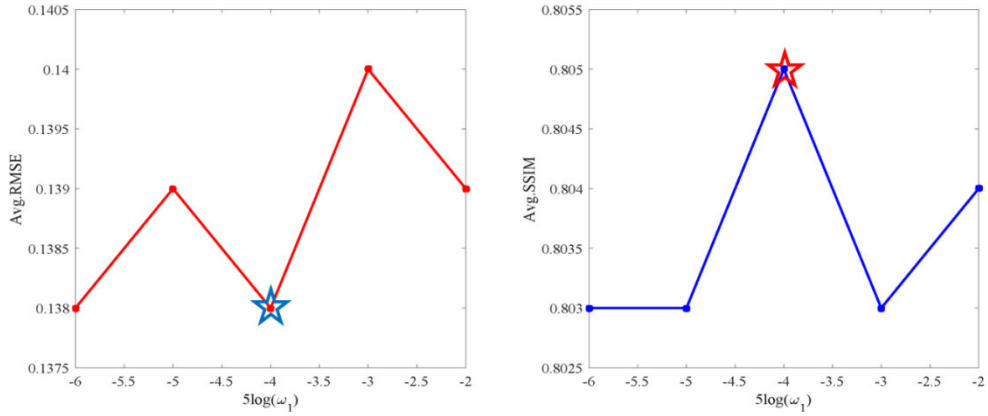


FIGURE 6. Plots of the average RMSEs (left) and average SSIMs (right) on the simulation test data by changing the parameter w_1 with 0.05, 0.005, 0.0005, 0.00005, and 0.000005 when the w_2 was 0.075, the δ was 10^{-6}. The parameter w_1 chosen in the manuscript was marked by ☆.

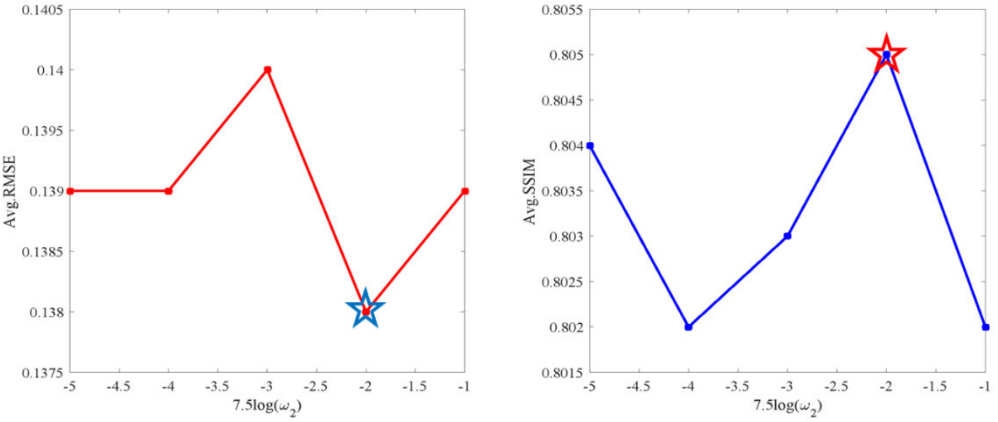


FIGURE 7. Plots of the average RMSEs (left) and average SSIMs (right) on the simulation test data by changing the parameter w_2 with 0.75, 0.075, 0.0075, 0.00075, 0.000075 when the w_1 was 0.0005, the δ was 10^{-6}. The parameter w_2 chosen in the manuscript was marked by ☆.

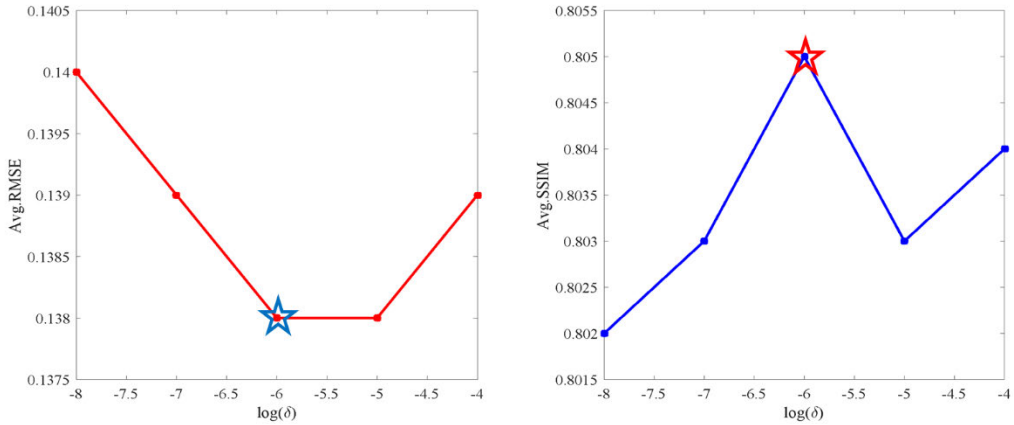


FIGURE 8. Plots of the average RMSEs (left) and average SSIMs (right) on the simulation test data by changing the parameter δ with 10^{-4}, 10^{-5}, 10^{-6}, 10^{-7}, 10^{-8} when the w_1 was set to be 0.0005, the w_2 was 0.075. The parameter δ chosen in the manuscript was marked by ☆.

To further quantify the performance of denoising methods. By increasing the MLEM iteration number with 24, 48, 72, 96, and 120, the CRC of gray matter versus STD in white matter curves of slice 34 were plotted in Fig. 11. Compared

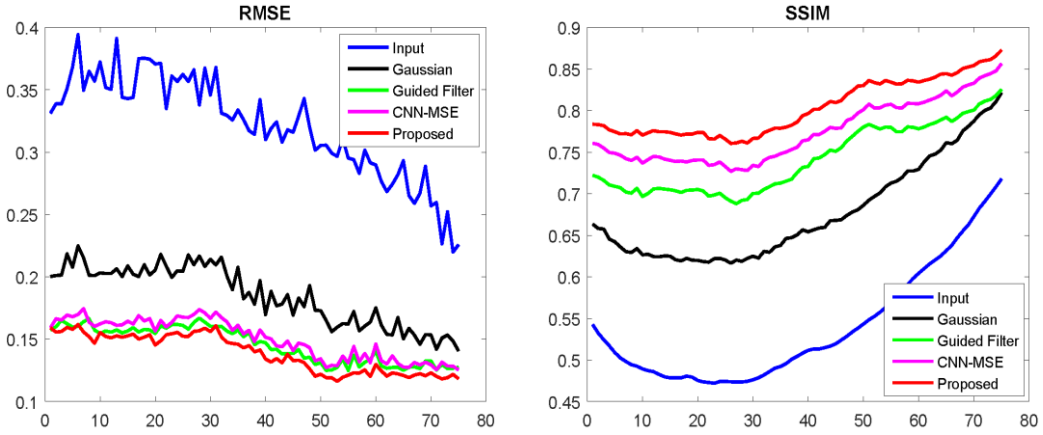


FIGURE 9. The RMSE and SSIM of each test resulting image (a total of 75 images) using different denoising algorithms (Gaussian filter, guided filter, CNN-MSE, and the proposed method).

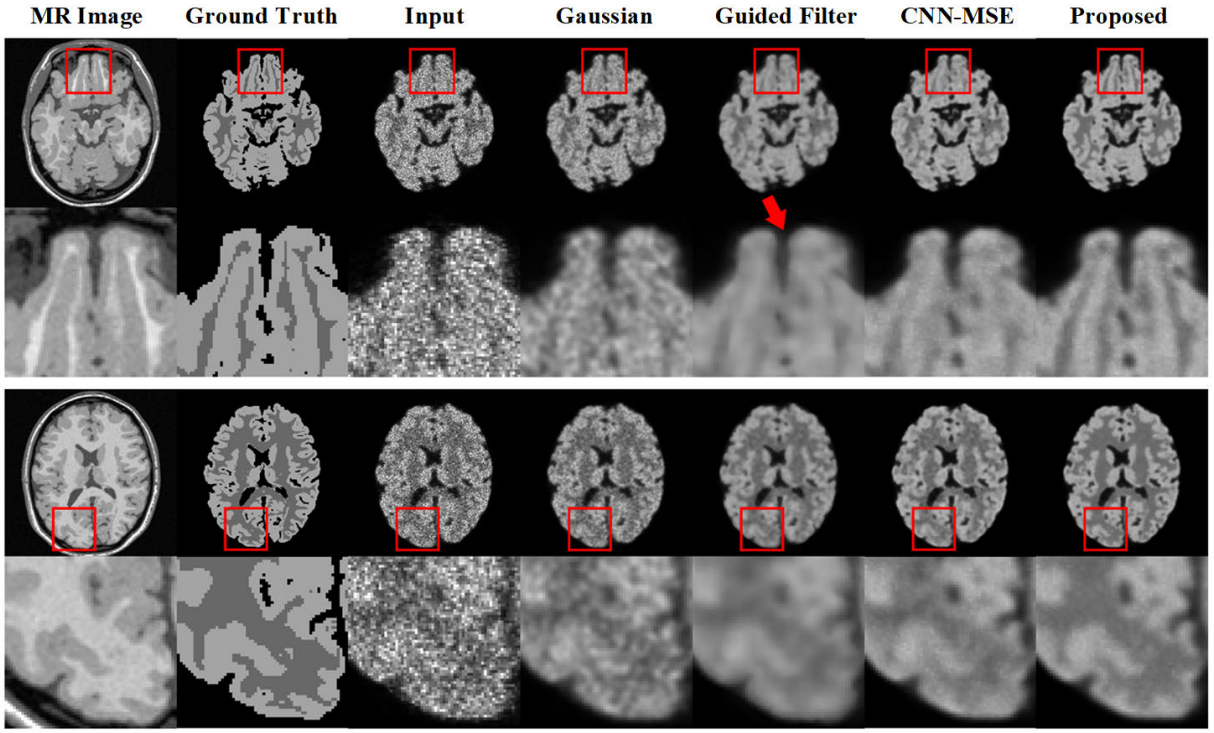


FIGURE 10. The transverse view of the slice 1 (up) and slice 34 (bottom) from test phantom processed by Gaussian filter, guided filter, CNN trained using MSE loss, and the proposed method. Two cortex regions from the transverse view are zoomed in for easier visual comparison.

with other methods, the proposed method performs the highest CRC at any matched STD level. The overall ensemble variance versus means squared bias curves by varying the MLEM iteration number were also plotted in Fig. 12. Results show that the proposed method can achieve less bias at a fixed variance than number were also plotted in Fig. 12. Results show that the proposed method can achieve less bias at a fixed variance than others while the ensemble variance keeps low.

Then, we compared the performance of each method by varying the corresponding parameters. The guided filter changed the radius r ranged from 1 to 5 and the regularization parameter ε was fixed to 10^{-2}. The Gaussian filter changed the FWHM ranged from 1.5 pixels to 5.5 pixels, and the CNN denoising methods changed the number of training iterations with 20 epochs, 60 epochs, 100 epochs, 150 epochs, and 200 epochs. The number of MLEM reconstruction iterations was fixed at 120. The noisy input was generated by MLEM changed the reconstruction iterations with 24, 48, 72, 96, and 120. The CRC of gray matter versus STD in white matter curves of slice 34 by varying the corresponding parameters were plotted in Fig. 13. The proposed method can achieve the highest CRC level at any matched STD than other denoising

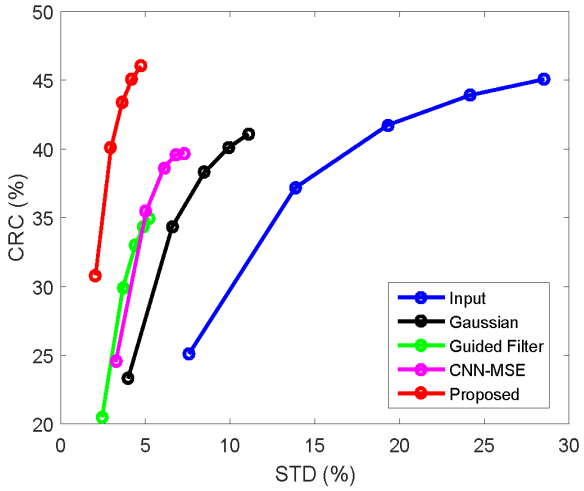


FIGURE 11. Plot of the CRC versus STD trade-off curves of slice 34 (generated with increasing MLEM iterations of 24, 48, 72, 96, and 120) obtained using: (i) Input. (ii) Gaussian, (iii) Guided Filter, (iv) CNN-MSE, and (v) Proposed.

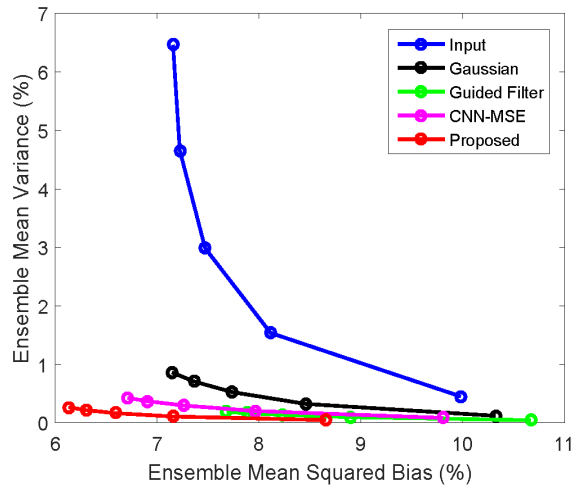


FIGURE 12. Plot of the ensemble variance versus ensemble means squared bias trade-off curves of slice 34 (generated with increasing MLEM iterations of 24, 48, 72, 96, and 120) obtained by using: (i) Input. (ii) Gaussian, (iii) Guided Filter, (iv) CNN-MSE, and (v) Proposed.

methods while the STD keeps low. With the increasing radius of the filter window, the guided filter can obtain slightly less background noise than the proposed method, but the contrast between gray matter and the white matter is inferior. Fig. 14 shows the ensemble variance versus means squared bias curves, which indicates that the proposed method can achieve less bias at a fixed variance than others while the ensemble variance keeps low.

To provide a visual evaluation, Fig. 15 shows the images with matching CRC level labeled by star markers in Fig. 13. It is clearly seen that the proposed method can reduce the noise level than others. The RMSE of the input image, images processed by Gaussian filter, guided filter, CNN-MSE, and the proposed method are 0.321, 0.262, 0.188, 0.165, and 0.159 respectively.

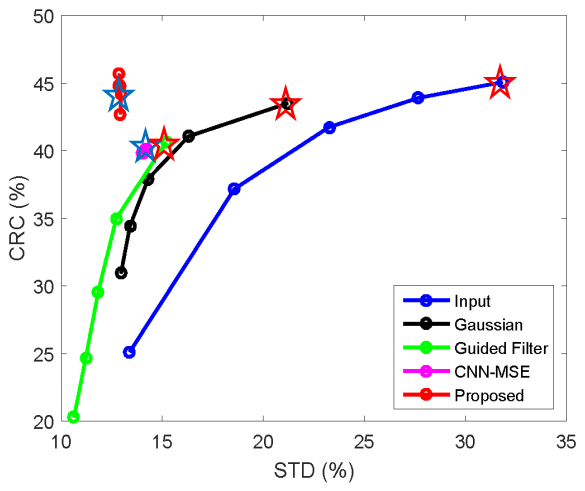


FIGURE 13. The CRC between the gray matter and the white matter versus the STD in white matter curves of slice 34 by varying the main parameters of each method, for (i) Input (increasing iterations), (ii) Gaussian (increasing FWHM), (iii) Guided Filter (increasing windows size), (iv) CNN-MSE (increasing trained iterations), and (v) Proposed (increasing trained iterations). The point with similar CRC value labeled by star markers will be used for visual evaluation in Fig. 15.

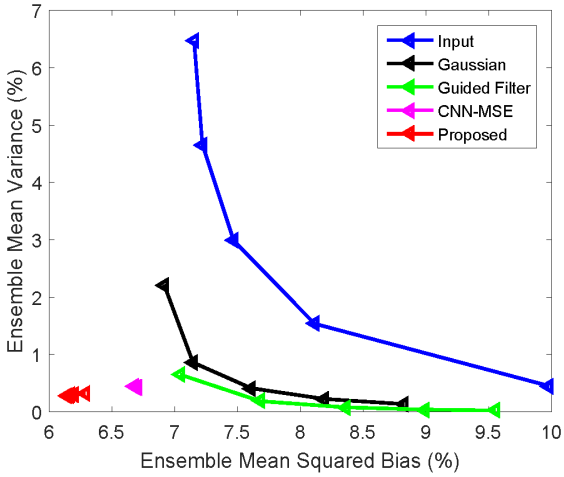


FIGURE 14. Plot of the ensemble variance versus ensemble means squared bias trade-off curves of slice 34 by varying the main parameters of each method, for (i) Input (increasing iterations), (ii) Gaussian (increasing FWHM), (iii) Guided Filter (increasing windows size), (iv) CNN-MSE (increasing trained iterations), and (v) Proposed (increasing trained iterations).

B. PATIENT STUDY

Following extensive simulation experiments and evaluations, we further confirmed the effectiveness of the proposed method on the clinical dataset. Due to the true activities values of clinical data are unknown, we cannot adjust the parameters of methods based on the evaluated metrics as used in simulated experiments. To balance the contrast and noise in the denoised image, we empirically set the window radius and the regularization parameter in the guided filter to 1 and 10^{-2}, respectively, while the FWHM of the Gaussian filter was set to 1.5 pixels. Fig. 16 shows the images

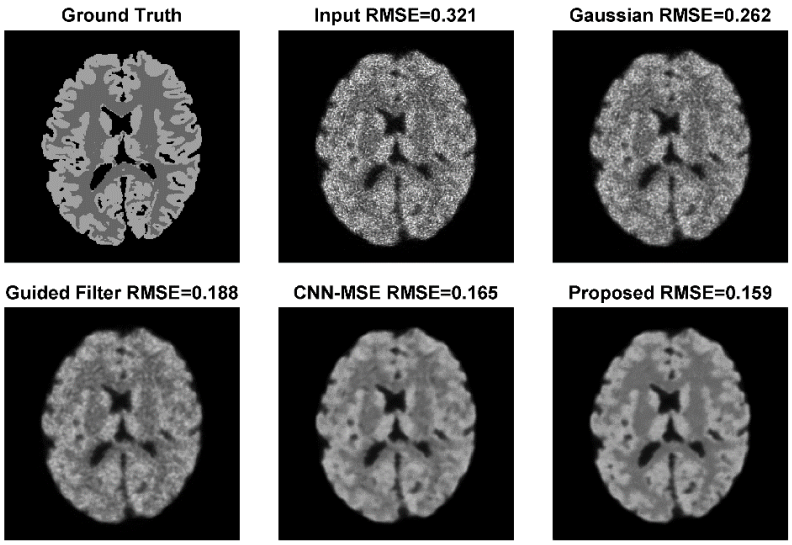


FIGURE 15. The denoised images using different methods with different parameters (Gaussian: FWHM = 1.5 pixels; Guided Filter: window radius = 1 pixel; input: iteration = 120; CNN-MSE: 20 epochs; proposed: 60 epochs). The images were selected by matching the CRC level approximately (see Fig. 13).

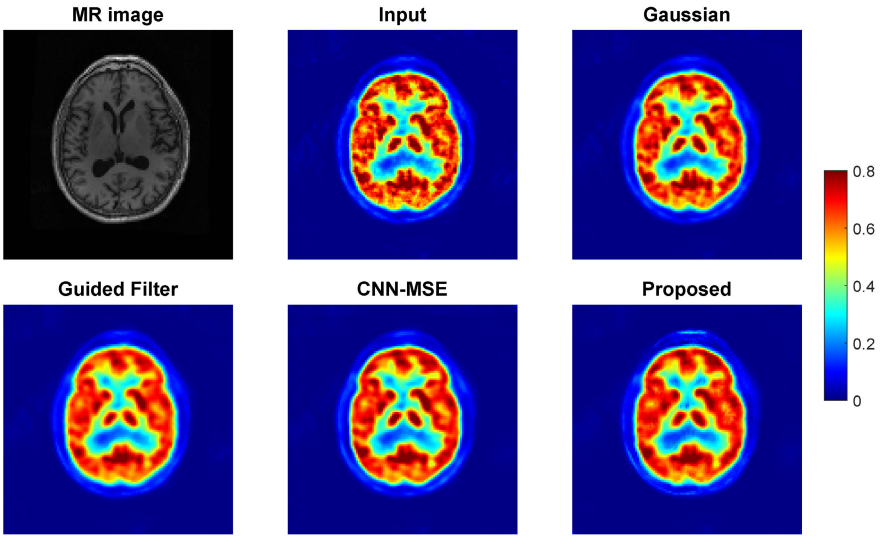


FIGURE 16. The transverse views of slice 16 of clinical brain test data denoising with different methods. The parameters was chosen to balance the noise and contrast (Gaussian: FWHM = 1.5 pixels; Guided Filter: r = 1, $\varepsilon = 10^{-2}$; CNN-MSE and Proposed: 50 epochs).

for visual evaluation processed by Gaussian filter, guided filter, CNN-MSE, and the proposed method. We can see that all methods can effectively reduce noise, but the Gaussian filter and guided filter result more blurry than CNN and proposed methods. To provide the quantitative comparison of the denoised images, we randomly selected twenty square ROIs with the size of 12 × 12 mm^2, and then calculated the mean activity and STD in ROIs, respectively. Fig. 17 shows the regional noise versus mean activity curves generated by increasing the iteration number of MLEM. We can see that the proposed method can achieve lower STD with matched mean activity compared with Gaussian filter. The CNN-MSE and guided filter reduce the STD at the expense of increasing mean activity bias.

We further compared the mean activity versus STD curves of the twenty square ROIs with the size of 12 × 12 mm^2 (Fig. 17) with the curves of more ROIs (Fig. 18) and the larger ROIs (Fig. 19). The more ROIs were generated by randomly adding ten ROIs with the size of 12 × 12 mm^2 based on the twenty ROIs in Fig. 17. The larger ROIs were generated by enlarging the ROI size from 12 × 12 mm^2 to 14 × 14 mm^2 while the positions and number were consistent with the ROI in Fig. 17. It is known that MLEM with Poisson likelihood is asymptotically unbiased. Thus, the bias of the

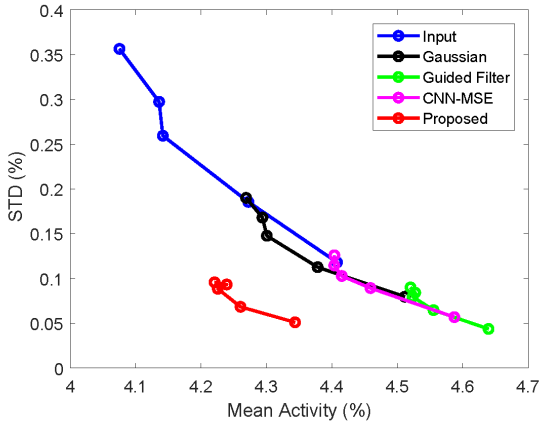


FIGURE 17. Plot of the STD versus Mean Activity trade-off curves of slice 16 of clinical brain test data changing with the iteration number (increasing iteration of 24, 48, 72, 96, and 120) obtained using: (i) Input, (ii) Gaussian, (iii) Guided Filter, (iv) CNN-MSE, and (v) Proposed. Twenty ROIs with the size of 12 × 12 mm^2 were randomly selected.

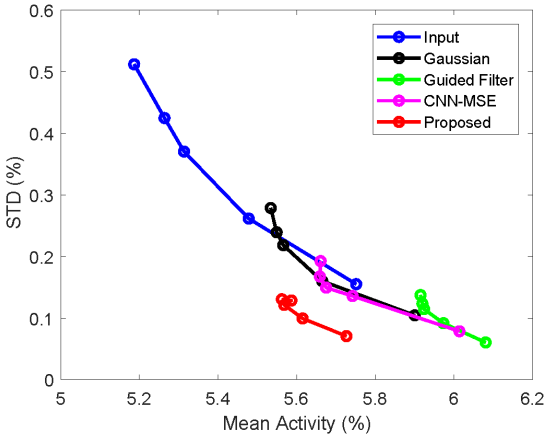


FIGURE 18. Plot of the STD versus Mean Activity trade-off curves of slice 16 of clinical brain test data changing with the iteration number. Thirty square ROIs with the size of 12 × 12 mm^2 were selected, as generated by adding ten ROIs randomly based on the twenty ROIs from Fig. 17.

denoised method can be evaluated roughly combined with the mean activity of MLEM (Input, labeled with the blue curves). As can be seen in Fig.17-18, the proposed method can achieve the lowest STD at the matched mean activity than others and the bias keep low whether the more ROIs or the larger ROIs were selected.

V. DISCUSSION

A. SELECTION OF THE LOSS FUNCTION

The loss function is very critical for the network, while the commonly used loss function MSE usually results in blurred outputs. Several solutions have been proposed to reduce the blurry, such as adding gradient terms [22], [30], using perceptual loss [23], SSIM loss [31], and L1 loss [15]. Gong *et al.* proposed perceptual loss based on L2 norm which compared feature maps instead of pixel intensities [23]. The loss function needs the additional network and the assumption that low-level features trained with natural images are also present in medical images is worth explaining. Nie *et al.*

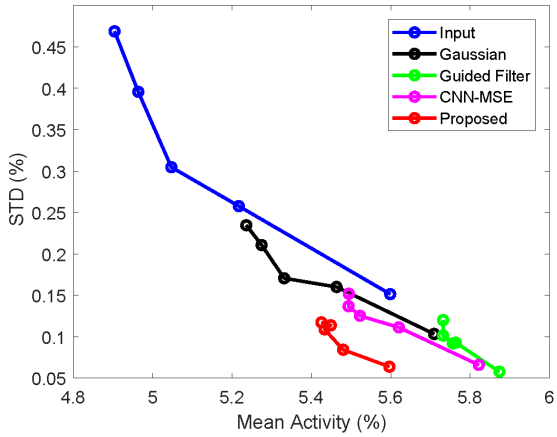


FIGURE 19. Plot of the STD versus Mean Activity trade-off curves of slice 16 of clinical brain test data changing with the iteration number. Twenty square ROIs with the size of 14 × 14 mm^2 were selected, as generated by enlarging the size of ROIs from Fig. 17 while the positions and number were fixed.

proposed loss function includes MSE, binary cross-entropy loss, and image gradient loss which is consistent with the third term of our proposed loss function [30]. Kaplan *et al.* proposed to add gradient and total variation with the L2 norm loss which are the same as the second and third terms of our proposed [22]. The above approaches can reduce the blurry from MSE loss but cannot avoid it. The cost function of the guided filter is based on MSE which may be one of the reasons for the blurry outputs, as shown in Fig. 10. Moreover, the use of MSE as the loss function is under the assumption that the outputs are corrupted by additive white Gaussian noise, which is not applicable for low-count PET images that suffer from the complex noise model. In this work, we used the L1-norm and gradient terms instead of the MSE, as shown in Fig. 11 and Fig. 15, the proposed loss has better noise reduction at matching contrast level no matter changed the iteration number or the parameters. This proves that the proposed loss is more suitable for low-count PET denoising than MSE. Due to the gradient terms, our proposed can preserve the image edge and keep the details close to the label image as shown in the rectangular red box of Fig. 10.

B. SELECTION OF THE NETWORK ARCHITECTURE

Our proposed network consists of a cascade of convolution layers and ReLU except for the final convolution layer, without any pooling and BN layers. The main reason is that pooling operations down-sampling the dimension of feature maps are usually used in classification and recognition tasks but not suitable for pixel-wise tasks like PET image denoising. The BN layers which using the statistic of the training dataset during testing are more likely to introduce artifacts and limit the generalization ability. Recently, abundant advanced networks were used for PET denoising. The 2D cycle Wasserstein regression adversarial networks (CycleWGANs) [20] and the 3D conditional regression adversarial networks (3D c-GANs) [32] were proposed to boost low-dose PET image

TABLE 2. Quantitative evaluations on proposed method with different number of layers. The average RMSEs and average SSIMs were calculated on the test dataset.

	Avg. RMSEs	Avg. SSIMs
12 layers	0.138 ± 0.016	0.805 ± 0.034
11 layers	0.139±0.016	0.804±0.034
10 layers	0.139±0.016	0.803±0.034
9 layers	0.141±0.016	0.803±0.034
8 layers	0.140±0.016	0.802±0.034

Values are expressed as means ± standard deviations.

quality, where the generators include block residual modules and 3D U-Net-like network respectively. The encoder-decoder CNN with U-Net structure was proposed to reduce the image noise from ultra-low-dose PET data [33]. However, the frameworks with skip connection were applied to end-to-end studies well as mentioned above, which learn the directed mapping between low-count and high-count PET images.

In this work, we learned the linear representation coefficients of SVLRM instead of the end-to-end trainable network to generate high-quality PET images. The proposed deep learning-based joint filter introduced the MR images as the prior information. By the supervision of the image label, the linear representation coefficients can determine well whether the structures of the MR image should be transferred to the denoised PET image. Moreover, with the pixel-wise learning of the linear coefficients in the proposed method instead of the average operation on overlapping windows in the guided filter, high-frequency information can be preserved better.

To further prove the simplicity and effectiveness of the network's structure in this work, we compared the denoised performance of the proposed method with the different number of layers, as shown in TABLE 2. We can see that the CNN with 12 layers has the lowest average RMSE and the highest average SSIM than with fewer layers. Thus, the structure of the network proposed here is minimalist and can achieve the lowest complexity for a level of efficiency and quality output. Moreover, the patch size also affects the denoising performance of network, which should be larger than the receptive field. In this work, we used the CNN with 12 layers, of which the convolution kernel is 3 × 3 pixels and the stride value is 1. Thus, the receptive field of our network is 25 and the patch size of 64 × 64 was reasonable. We further compared the training losses with different patch sizes (32 × 32, 64 × 64, and 96 × 96), as shown in Fig. 20. We can see that the smaller patch size can acquire less loss at the end of training and learn the finer image detail. However, as shown in TABLE 3, the patch size of 64 × 64 can achieve the lowest average RMSE on the simulated test data than others. It can be explained that the appropriate larger patch size is necessary to require more contextual information when the input image

TABLE 3. Denoising performance (in average RMSE) on the simulated test data for different sizes of training input patches, keeping all other parameters constants. The testing inputs were the whole PET and MR images.

Training patch size	32 × 32	64 × 64	96 × 96
Avg. RMSEs	0.146 ± 0.015	0.138 ± 0.016	0.142 ± 0.015

Values are expressed as means ± standard deviations.

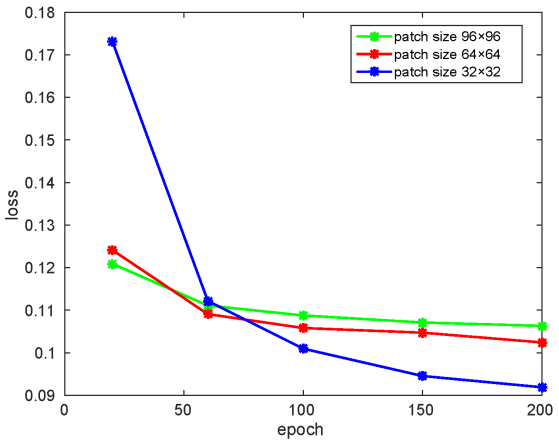


FIGURE 20. Plot of the training losses generated with changing the number of training epochs of 20, 60, 100, 150, and 200 with different patch sizes (32 × 32, 64 × 64, and 96 × 96).

with higher noise level. Thus, we used the path size of 64 × 64 during training.

C. SELECTION OF THE EXPERIMENT DATASET

The purpose of our work is to denoising the last frame of dynamic PET image through a supervised framework, i.e., estimating the high-quality standard-count PET (HPET) image from low-quality low-count PET (LPET) image. We combined the long-scanned PET data to generate the HPET and down-sampling it to obtain LPET. The data acquisitions of [32] and [34] were both conducted twice (short and long periods) to obtain the LPET and HPET data, which would cause unnecessary harm to patients undoubtedly. Zhou *et al*. obtained the LPET image by down-sampling the HPET raw data randomly, which is similar to ours but they need a large number of patients for training [20]. In this work, only a single dynamic scan is demanded for each patient which avoids additional harm. Due to fine-tuning, vast clinical data are not required for training.

The ideal PET system without any attenuation and object-dependent scatter was assumed in this work [21]. We further compared the quantitative results including attenuation or not. The attenuation map (Fig. 21 (b)) was generated using the discrete MR images (Fig. 21 (a)) 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 attenuation map was incorporated into the forward projection to generate the realistic simulation PET sinogram data and also incorporated into

TABLE 4. Quantitative evaluation of all simulated test datasets which including attenuation or not in terms of the average RMSEs and average SSIMs.

	Avg. RMSEs	Avg. RMSEs (with attenuation)	Avg. SSIMs	Avg. SSIMs (with attenuation)
Input	0.322 ± 0.040	0.322±0.037	0.510 ± 0.070	0.540±0.07
Gaussian	0.185 ± 0.024	0.189±0.023	0.677 ± 0.060	0.699±0.52
Guided Filter	0.149 ± 0.014	0.152±0.015	0.741 ± 0.041	0.723±0.042
CNN-MSE	0.150 ± 0.016	0.158±0.016	0.774 ± 0.037	0.748±0.039
Proposed	**0.138 ± 0.016**	**0.148±0.016**	**0.805 ± 0.034**	**0.772±0.036**

Values are expressed as means ± standard deviations.

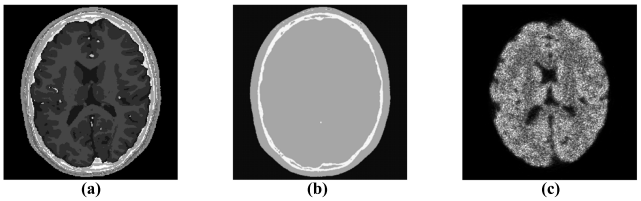


FIGURE 21. Transaxial slices based on BrainWeb used for simulation from the same subject. (a) the discrete MR image; (b) the attenuation map; (c) PET noisy input without attenuation; (d) PET noisy input with attenuation.

maximum-likelihood expectation-maximization (MLEM) reconstruction. The PET noisy image with attenuation was shown in the Fig. 21 (c). We compared the average RMSEs and average SSIMs of the all simulated test dataset including attenuation or not, as shown in TABLE 4. It can be seen that the proposed method has the lowest average RMSE and the highest average SSIM no matter include attenuation or not. The attenuation affects the PET noisy input, but have little effect on the trend of comparison results denoised by different post-filters. The scatter was difficult for analytic simulation which was not evaluated. As the object-dependent scatter only increases the noise level, it may similar to attenuation and will have little effect on the trend of comparison results denoised by different post-filters.

The last frame of the dynamic PET image was used as testing input, instead of the down-sampled PET image, to prove the generalization performance of the proposed. It's worth mentioning that our network can also denoise the static images, of which noise level is similar to the testing inputs. Besides, other frames of dynamic PET data can be denoised with the same framework by changing the down-sampled degree of training input data.

D. LIMITATION AND FURTHER WORK

The proposed method can reduce noise while preserving the structures in Fig. 10, but this is not very clear by visual evaluation in the results of clinical experiments as shown in Fig. 16. The possible reasons are as follows. 1) The label images are blurred in training. The label generated by summing the entire dynamic PET data into one frame usually lost image details though reduced noise. 2) Poor registration of clinical PET/MR images. In the simulation study, the structures of the PET image are consistent with the MR image for the same brain phantom, thus the structures can be preserved when the MR image was used as guidance. As the clinical PET/MR data were acquired from different machines at different months, the slice thickness and image array in PET images and MR images are different. The data acquired from integrated PET/MR systems or finer registration methods may improve the quality of the resulting image. Nevertheless, the proposed method can obtain the lowest noise level compared with other methods for a matched activity value in Fig. 17-19. It confirms the superiority of the proposed method in the clinical study. Due to the technique that pre-training with simulated data followed by fine-tuning only the last two layers with clinical data, the network will be quickly and effectively trained when the fine registered PET/MR images were obtained in the future.

In our network, 2D convolution was used, thus the axial information was not extracted. We will extend the network to 3D convolution in the future. The clinical dataset acquired from integrated PET/MR scanners is expected to validate the performance of our method further, and more evaluations for clinical data will be explored in our future work.

VI. CONCLUSION

In this study, we proposed a deep learning-based joint filtering to improve the image quality of the last frame in dynamic PET scanning. The L1-norm was combined with edge-preserving and structure-preserving features as the loss function in training. We pre-trained the network using digital phantoms and then fine-tuned the last two convolution layers of the network using real brain data. The simulation and clinical experiments show that the result images processed by the proposed method can reduce noise better than the Gaussian, guided filter, and the CNN trained using MSE loss function.

REFERENCES

[1] M. Bentourkia and H. Zaidi, "Tracer kinetic modeling in PET," *PET Clinics*, vol. 2, no. 2, pp. 267–277, Apr. 2007.
[2] A. J. Reader and H. Zaidi, "Advances in PET image reconstruction," *PET Clinics*, vol. 2, no. 2, pp. 173–190, Apr. 2007.

[3] L. Lu, N. A. Karakatsanis, J. Tang, W. Chen, and A. Rahmim, "3.5 D dynamic PET image reconstruction incorporating kinetics-based clusters," *Phys. Med. Biol.*, vol. 57, no. 15, p. 5035, Jul. 2012.

[4] 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.

[5] 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.

[6] 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.

[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] Z. Xu, M. Gao, G. Z. Papadakis, B. Luna, S. Jain, D. J. Mollura, and U. Bagci, "Joint solution for PET image segmentation, denoising, and partial volume correction," *Med. Image Anal.*, vol. 46, pp. 229–243, May 2018.

[9] C. Tomasi and R. Manduchi, "Bilateral filtering for gray and color images," in *Proc. 6th Int. Conf. Comput. Vis.*, Jan. 1998, pp. 839–846.

[10] F. Hofheinz *et al.*, "Suitability of bilateral filtering for edge-preserving noise reduction in PET," *EJNMMI Res.*, vol. 1, no. 1, pp. 1–9, Oct. 2011.

[11] K. He, J. Sun, and X. Tang, "Guided image filtering," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 35, no. 6, pp. 1397–1409, Jun. 2012.

[12] L. Lu, D. Hu, X. Ma, J. Ma, A. Rahmim, and W. Chen, "Dynamic PET denoising incorporating a composite image guided filter," in *Proc. IEEE Nucl. Sci. Symp. Med. Imag. Conf. (NSS/MIC)*, Nov. 2014, pp. 1–4.

[13] 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.

[14] J. Yan, J. C. S. Lim, and D. W. Townsend, "MRI-guided brain PET image filtering and partial volume correction," *Phys. Med. Biol.*, vol. 60, no. 3, p. 961, Jan. 2015.

[15] J. Pan, J. Dong, J. S. Ren, L. Lin, J. Tang, and M.-H. Yang, "Spatially variant linear representation models for joint filtering," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2019, pp. 1702–1711.

[16] L. Bertinetto, J. Valmadre, J. F. Henriques, A. Vedaldi, and P. H. S. Torr, "Fully-convolutional siamese networks for object tracking," in *Proc. Eur. Conf. Comput. Vis.*, Oct. 2016, pp. 850–865.

[17] J. Long, E. Shelhamer, and T. Darrell, "Fully convolutional networks for semantic segmentation," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2015, pp. 3431–3440.

[18] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet classification with deep convolutional neural networks," in *Proc. Adv. Neural Inf. Process. Syst. (NIPS)*, 2012, pp. 1097–1105.

[19] 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, Aug. 2019.

[20] L. Zhou, J. D. Schaefferkoetter, I. W. K. Tham, G. Huang, and J. Yan, "Supervised learning with cyclegan for low-dose FDG PET image denoising," *Med. Image Anal.*, vol. 65, Oct. 2020, Art. no. 101770.

[21] 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.

[22] S. Kaplan and Y.-M. Zhu, "Full-dose PET image estimation from low-dose PET image using deep learning: A pilot study," *J. Digit. Imag.*, vol. 32, no. 5, pp. 773–778, Oct. 2019.

[23] 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.

[24] R. K.-S. Kwan, A. C. Evans, and G. B. Pike, "MRI simulation-based evaluation of image-processing and classification methods," *IEEE Trans. Med. Imag.*, vol. 18, no. 11, pp. 1085–1097, Nov. 1999. [Online]. Available: http://brainweb.bic.mni.mcgill.ca/brainweb/

[25] B. Aubert-Broche, M. Griffin, G. B. Pike, A. C. Evans, and D. L. Collins, "Twenty new digital brain phantoms for creation of validation image data bases," *IEEE Trans. Med. Imag.*, vol. 25, no. 11, pp. 1410–1416, Nov. 2006.

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

[27] J. A. Fessler. (2010). *Image Reconstruction Toolbox (MATLAB)*. [Online]. Available: http://web.eecs.umich.edu/~fessler/irt/irt

[28] C. R. Jack *et al.*, "The Alzheimer's disease neuroimaging initiative (ADNI): MRI methods," *J. Magn. Reson. Imag.*, vol. 27, no. 4, pp. 685–691, 2008. [Online]. Available: http://adni.loni.usc.edu/

[29] S. Pieper, M. Halle, and R. Kikinis, "3D slicer," in *Proc. 2nd IEEE Int. Symp. Biomed. Imaging, Macro Nano*, Apr. 2004, pp. 632–635.

[30] D. Nie *et al.*, "Medical image synthesis with context-aware generative adversarial networks," in *Proc. Int. Conf. Med. Image Comput. Comput. Assist. Interv.*, Quebec City, QC, Canada, 2017, pp. 417–425.

[31] J. Snell, K. Ridgeway, R. Liao, B. D. Roads, M. C. Mozer, and R. S. Zemel, "Learning to generate images with perceptual similarity metrics," in *Proc. IEEE Int. Conf. Image Process. (ICIP)*, Sep. 2017, pp. 4277–4281.

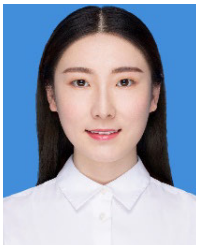
[32] Y. Wang, B. Yu, L. Wang, C. Zu, D. S. Lalush, W. Lin, X. Wu, J. Zhou, D. Shen, and L. Zhou, "3D conditional generative adversarial networks for high-quality PET image estimation at low dose," *NeuroImage*, vol. 174, pp. 550–562, Jul. 2018.

[33] K. T. Chen *et al.*, "Ultra-low-dose 18F-florbetaben amyloid PET imaging using deep learning with multi-contrast MRI inputs," *Radiology*, vol. 290, no. 3, pp. 649–656, 2019.

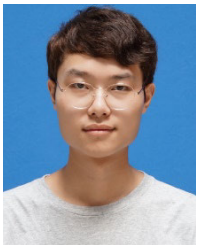
[34] L. Xiang, Y. Qiao, D. Nie, L. An, W. Lin, Q. Wang, and D. Shen, "Deep auto-context convolutional neural networks for standard-dose PET image estimation from low-dose PET/MRI," *Neurocomputing*, vol. 267, pp. 406–416, Dec. 2017.

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.



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.



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 interest includes PET imaging.

FANGHU WANG received the M.Eng. degree in biomedical engineering from Southern Medical University, in 2020. He is currently an Engineer with the WeiLun PET Center, Department of Nuclear Medicine, Guangdong Provincial People's Hospital and Guangdong Academy of Medical Sciences. His research interests include cardiac PET imaging and analysis.

WENBING LV received the Ph.D. degree in biomedical engineering from Southern Medical University, China, in 2020. She is currently a Postdoctoral Research Fellow with the Guangdong Provincial Key Laboratory of Medical Image Processing, School of Biomedical Engineering, Southern Medical University. Her research interests include PET image analysis and radiomics.

HUOBIAO ZHU received the B.Eng. degree from Southern Medical University, in 2018, where he is currently pursuing the master's degree in engineering with the Guangdong Provincial Key Laboratory of Medical Image Processing, School of Biomedical Engineering. His research interest includes PET reconstruction.

LIJUN LU received the Ph.D. degree in biomedical engineering from Southern Medical University, in 2012. He is currently a 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.