IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received June 29, 2019, accepted July 13, 2019, date of publication July 16, 2019, date of current version August 5, 2019.

Digital Object Identifier 10.1109/ACCESS.2019.2929230

Dynamic PET Image Denoising Using Deep Convolutional Neural Networks Without Prior Training Datasets

FUMIO HASHIMOTO[1], HIROYUKI OHBA[1], KIBO OTE[1], ATSUSHI TERAMOTO[2], AND HIDEO TSUKADA[1]

[1]Central Research Laboratory, Hamamatsu Photonics K.K., Shizuoka 434-8601, Japan
[2]School of Medical Sciences, Fujita Health University, Toyoake 470-1192, Japan

Corresponding author: Fumio Hashimoto (fumio.hashimoto@crl.hpk.co.jp)

This work was supported by the Hamamatsu Photonics K.K., Shizuoka, Japan.

ABSTRACT Deep learning has attracted growing interest for application to medical imaging, such as positron emission tomography (PET), due to its excellent performance. Convolutional neural networks (CNNs), a facet of deep learning requires large training-image datasets. This presents a challenge in a clinical setting because it is difficult to prepare large, high-quality patient-related datasets. Recently, the deep image prior (DIP) approach has been devised, based on the fact that CNN structures have the intrinsic ability to solve inverse problems such as denoising without pre-training and do not require the preparation of training datasets. Herein, we proposed the dynamic PET image denoising using a DIP approach, with the PET data itself being used to reduce the statistical image noise. Static PET data were acquired for input to the network, with the dynamic PET images being handled as training labels, while the denoised dynamic PET images were represented by the network output. We applied the proposed DIP method to computer simulations and also to real data acquired from a living monkey brain with ^{18}F-fluoro-2-deoxy-D-glucose (^{18}F-FDG). As a simulation result, our DIP method produced less noisy and more accurate dynamic images than the other algorithms. Moreover, using real data, the DIP method was found to perform better than other types of post-denoising method in terms of contrast-to-noise ratio, and also maintain the contrast-to-noise ratio when resampling the list data to 1/5 and 1/10 of the original size, demonstrating that the DIP method could be applied to low-dose PET imaging. These results indicated that the proposed DIP method provides a promising means of post-denoising for dynamic PET images.

INDEX TERMS Convolutional neural networks, deep image prior, deep learning, denoising, dynamic positron emission tomography.

I. INTRODUCTION

Positron emission tomography (PET) is a noninvasive imaging modality for dynamically measuring the pharmacokinetics of target-specific PET tracers in a living body. It is applied not only to cancer diagnosis but also to the early detection of neurodegenerative diseases such as Alzheimer's and Parkinson's disease [1]. To accelerate understanding of these diseases, brain-dedicated PET scanners which can perform dynamic measurements have been developed [2], [3]. In addition, animal PET scanners have also been developed for use in the development of novel PET tracers and brain function research [4].

To measure the kinetics and distribution of PET tracers dynamically, the use of short time frame data is indispensable, although this results in a high level of noise. Previously, a Gaussian filter (GF) has usually been applied for post-denoising to improve the signal-to-noise ratio, as well as to facilitate the identification of sufficiently large structures. However, a GF reduces the spatial resolution of the PET images. Therefore, various approaches to image reconstruction or post-denoising have been adopted for the denoising of PET images without impairing their quantitativeness. For example, maximum a posteriori image reconstruction incorporated with various prior models (i. e., edge preservation

The associate editor coordinating the review of this manuscript and approving it for publication was Shubhajit Roy Chowdhury.

priors [5]–[7], voxel kinetics priors [8], anatomical priors [9]–[11], and a kernel method [12] have been reported. However, their high computational costs have prevented their use in clinical applications. Therefore, post-denoising is performed separately from the image reconstruction process. Bian et al. developed a kinetics-induced bilateral filter extended in the time domain, and improved the quantitativeness of the dynamic PET images [13]. Furthermore, Dutta et al. applied a non-local means (NLM) filter [14] to dynamic PET images [15]. In addition, some post-denoising methods, such as HYPR processing [16], image guided filtering (IGF) [17], and wavelet denoising [18] have been developed. All exhibit a higher level of performance than the GF.

Deep learning has recently attracted increased interest in the medical imaging field, and offered better performance than conventional model-based algorithms in terms of denoising and segmentation tasks [19]–[23]. Generally, deep learning, such as that associated with convolutional neural networks (CNN), requires the preparation of large training-image datasets. This presents a challenge in a clinical setting because it would be very difficult to prepare large, high-quality patient-based datasets. When deep learning is applied to PET image denoising, it is necessary to prepare high-dose or long-time-frame and low-dose or short-time-frame image pairs, however, it is difficult to obtain a ''clean'' label for dynamic PET images. If images acquired from disease cases are not included in the training datasets, there are cases when the trained network cannot exactly denoise the PET images acquired from patients. In addition, training datasets for novel PET tracers are impossible to prepare.

Some recent reports have shown that CNN can produce a denoised image without ''clean'' dataset [24] or a training dataset [25], [26]. Particularly, the deep image prior (DIP) approach described in [25] suggests that CNN structures have an intrinsic ability to solve inverse problems such as denoising without any pre-training. The DIP approach iterates learning using a pair of random noise and corrupted image and a denoised image is obtained by the network output with moderate iterations. Therefore, this approach would not require the preparation of training datasets and only random noise would need to be prepared as the network input to denoise the image. Furthermore, inspired by this approach, some studies have shown that high quality PET images can be obtained using anatomical information such as magnetic resonance (MR) and computed tomography (CT) images as the network input, relative to the use of random noise as the network input [27], [28]. In other words, to obtain high-quality PET images without prior training datasets, it is necessary to prepare MR or CT images without misregistration by using the PET/MR or CT scanner. However, when using brain-dedicated and animal PET scanners without the MR or CT system described in [2]–[4], it is impossible to obtain MR or CT images without misregistration.

In the present study, we propose the dynamic brain PET image denoising using DIP approach. The main advantage of this method is that anatomical information is not necessary; instead, the original PET data itself is used to reduce the statistical image noise. Our proposed method is regarded as being a type of ''unsupervised'' learning because it does not require an ideal label. Compared with the 3D NLM filter [29], [30] and IGF [17], our proposed method exhibited superior performance visually and quantitatively when applied to both a computer simulation and a real data obtained from a living monkey brain using ^{18}F-fluoro-2-deoxy-D-glucose (^{18}F-FDG).

II. DEEP IMAGE PRIOR [25]

In general, a neural network needs to train on a large dataset of image pairs. On the other hand, the DIP approach maps the CNN onto a single degraded image. This is based only on the prior information included in the CNN structure itself. The network weights θ are shown as a parameterization of the denoised output x, as follows:

$$\theta^* = \arg\min_{\theta} \|x_0 - f(\theta|z)\|, \quad x^* = f\left(\theta^*|z\right), \tag{1}$$

where $\|\cdot\|$ is the L2 norm, f represents the neural network, the training label x_0 is the corrupted image, and the network input z is random noise. Network f is trained from scratch with only random noise z and corrupted image x_0. With this framework, overfitting is caused by minimization of the L2 loss. However, it can be avoided by regularization of the CNN structure with moderate iteration.

III. METHODOLOGY

A. PROPOSED METHOD

The overview of the proposed method is illustrated in Figure 1. In our proposed DIP method, dynamic PET images, x_0^i ($i = 1, 2, \ldots, T$) were used as the training labels and static PET image z was used as the network input. The static image acquires all the data from the start to the end of data acquisition. The filtered dynamic PET images x^* are obtained as follows:

$$\theta^{i*} = \arg\min_{\theta^i} \left\|x_0^i - f\left(\theta^i|z\right)\right\|, \quad x^{i*} = f\left(\theta^{i*}|z\right). \tag{2}$$

With the proposed method, we use a patient's own PET image as the network input instead of random noise as the network input. In other words, the proposed DIP method can train only using a single data pair of the static PET image as network input and the dynamic PET image as label.

B. NETWORK ARCHITECTURE

Various state-of-the-art CNN algorithms have been validated for image-to-image transformation. In the present study, we employed an architecture based on 3D U-net [31], as shown in Figure 2.

The architecture consists of an encoding path (left side) and a decoding path (right side). The encoding path conforms to the typical architecture of a CNN, consisting of the repeated application of two $3 \times 3 \times 3$ 3D convolution layers, each followed by a batch normalization (BN) and a leaky rectified

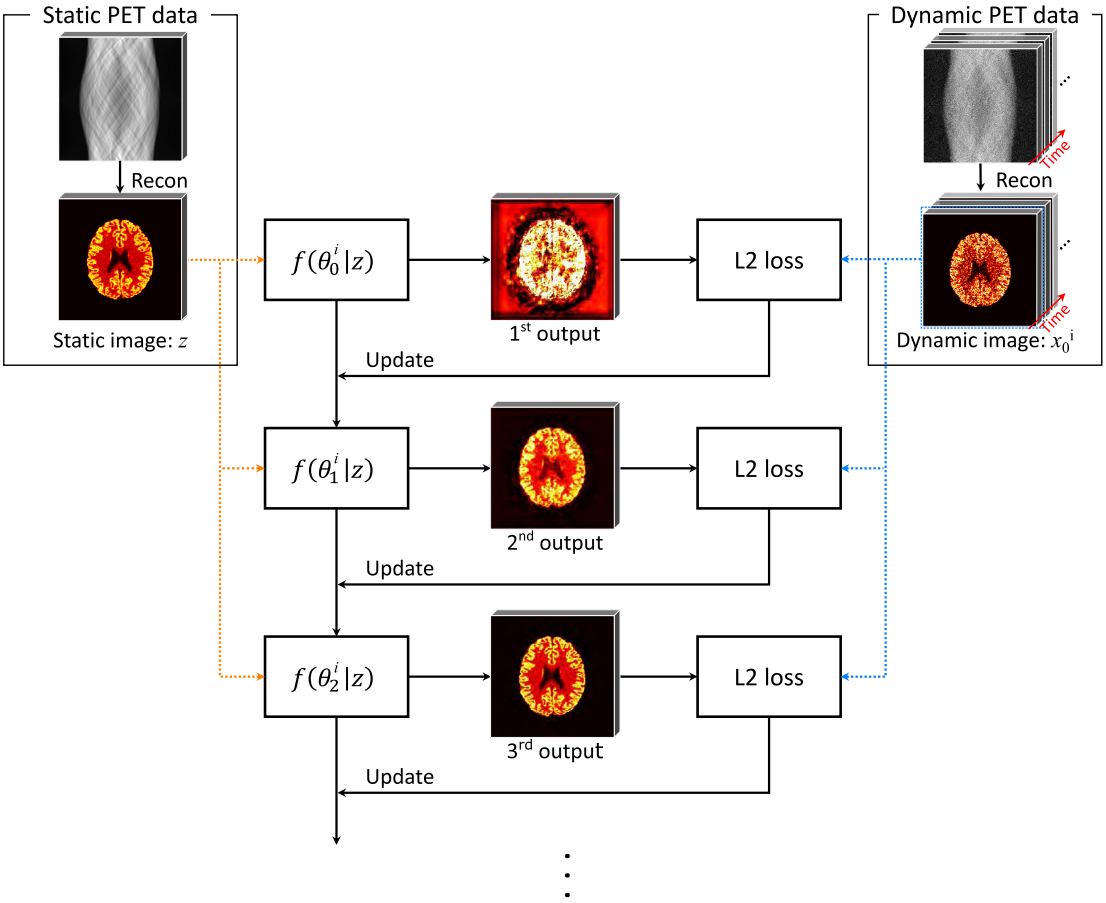

FIGURE 1. Schematic illustration of proposed DIP method. The dynamic PET images x_0 are used as training labels and the static PET image z is used as the CNN input. CNN training is started from a random weight θ_0, and θ are updated by minimizing the L2 loss. After training the CNN, filtered dynamic PET images are obtained as the final network output.

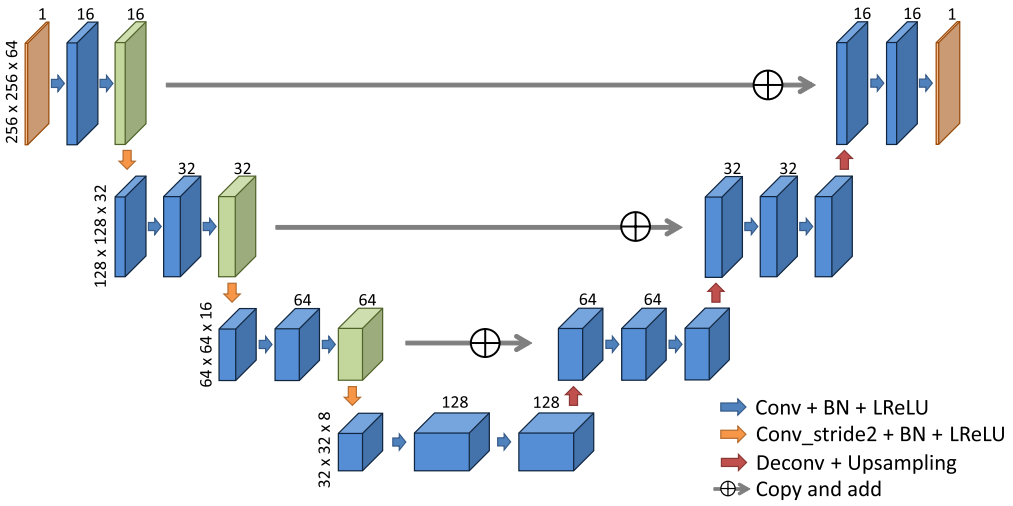

FIGURE 2. Architecture of neural network used for post-denoising. The number of channels is denoted at the top of each box. The pixel sizes appear at the left. The arrows denote the different operations.

linear unit (LReLU), as well as a $3 \times 3 \times 3$ 3D convolution layer with a stride of two for downsampling, followed by the BN and the LReLU. In addition, at each downsampling step, the number of feature channels is doubled. The decoding path consists of a $3 \times 3 \times 3$ deconvolution layer, followed by the BN and the LReLU, upsampling, skip connection with the

corresponding linked feature map from the encoding path, and two 3 × 3 × 3 3D convolution layers, each followed by the BN and the LReLU. Moreover, we activated the output layer through a linear function, and the mean squared error (MSE) was used as the loss function, as shown in (2). The stochastic gradient descent was used as the optimizer. To stabilize the network training, the input and label data were standardized by removing the mean and scaling to the unit variance. The network weights were initialized using He-initialization [32].

The U-net was run on a computer with Ubuntu 16.04, a graphic processing unit (NVIDIA Quadro P6000 with 24 GB of memory), Tensorflow 1.9 [33], and Keras 2.2.4 [34].

IV. EXPERIMENTAL SETUP

A. COMPUTER SIMULATION

We performed a computer simulation using a 3D brain phantom from BrainWeb [35], consisting of white and gray matter. A two-tissue compartment model was selected as the kinetic model, based on the glucose metabolism of ^{18}F-FDG, and the regional time activity curves (TACs) generated by Feng's method [36]. In our simulation, we used kinetic parameters K1, k2, k3 and k4, calculated from 13 healthy male volunteers in Table 1 [37]. Dynamic PET scanning commenced at the start of the bolus injection. The dynamic PET data consisted of 30 time frames over 90 min: 4 × 20 s, 4 × 40 s, 4 × 60 s, 4 × 180 s, and 14 × 300 s. The regional TACs were integrated in each time frame and dynamic PET sinograms were generated by simple forward projection. The acquired sinogram consisted of 256 bins, 256 angles, and 64 slices, while the reconstructed image had 256 × 256 × 64 voxels (1.0 × 1.0 × 2.0 mm per voxel). We assumed an ideal PET system without any scatter and attenuation to evaluate the simplified denoising performance of our proposed method. Poisson noise was added, resulting in the expected total number of coincidence counts, 10^9, for 90 min. Finally, the ordered subset-expectation maximization algorithm [38] with 6 iterations and 16 subsets was used for the reconstruction. Part of our dynamic PET simulation was implemented using scikit-image [39].

To evaluate the quantitative performance, we measured the peak signal-to-noise ratio (PSNR), structural similarity (SSIM) index [40], and the regional TACs of the gray and white matter. The PSNR is the power of the ratio between the maximum possible value of a signal and the MSE

$$\mathrm{PSNR} = 10\log_{10}\left(\frac{\max K^2}{\frac{1}{N_{\mathbf{R}}}\sum_{i\in\mathbf{R}}\left(K_i - K_i'\right)^2}\right), \tag{3}$$

where K and K' are the ground truth and the target image, $\mathbf{R}$ is the non-zero ground truth area, and $N_{\mathbf{R}}$ is the number of voxels in $\mathbf{R}$. The SSIM is a measure of the similarity between two images, which are proven to be consistent with human-eye perception

$$\mathrm{SSIM} = \frac{1}{N_{\mathbf{R}}}\sum_{i\in\mathbf{R}}\frac{\left(2\mu_{iK}\mu_{iK'}+c_1\right)\left(2\sigma_{iKK'}+c_2\right)}{\left(\mu_{iK}^2+\mu_{iK'}^2+c_1\right)\left(\sigma_{iK}^2+\sigma_{iK'}^2+c_2\right)}, \tag{4}$$

where μ_{iK}, $\mu_{iK'}$ and σ_{iK}, $\sigma_{iK'}$ are the average values and standard deviations in the square window centered on i-th voxel of images K and K', respectively. Meanwhile, $\sigma_{iKK'}$ is the covariance in the square window centered on the i-th voxel of images K and K', c_1 and c_2 are constants ($c_1 = (0.01L)^2$ and $c_2 = (0.03L)^2$). L is the dynamic range of the image. The square window size is 5 × 5.

B. PET MEASUREMENT FOR LIVING MONKEY BRAIN

All applicable international, national, and/or international guidelines for the care and use of animals were followed. Experiments were approved by the Ethical Committee of the Central Research Laboratory, Hamamatsu Photonics K.K. (Approval number HPK-2017-02). Dynamic PET measurements were performed on a rhesus monkey brain using an animal PET scanner (SHR-38000, Hamamatsu Photonics K.K., Japan). The monkey's head was rigidly fixed to the upper frame of a monkey chair using an acrylic head restraint, while the animal was conscious. The monkey, sitting in the restraining chair, was placed in the gantry in a fixed position, with the stereotactic coordinates aligned parallel with the orbitomeatal plane. Following a 30-min transmission scan using a ^{68}Ge-^{68}Ga rotation rod source, dynamic emission scanning was performed for 90 min after the injection of ^{18}F-FDG at a dose of 100.22 MBq. The data were reconstructed using a 3-D dynamic row-action maximum-likelihood algorithm [41] with 60 subsets and two iterations, while the attenuation was corrected using the transmission scan data. The dynamic PET data consisted of 30 time frames over 90 min: 4 × 20 s, 4 × 40 s, 4 × 60 s, 4 × 180 s, and 14 × 300 s. The individual PET and MR images were coregistered. The acquired sinogram consisted of 360 bins, 360 angles, and 103 slices with a span of three and a ring difference of 19, while the reconstructed image had 256 × 256 × 64 voxels (0.65 × 0.65 × 1.0167 mm/voxel). To evaluate the effect of the low-dose images, we prepared those low-dose images by resampling the list data to 1/5 and 1/10 of the size of the original.

For the evaluation, we measured the regional TACs of the gray and white matter and then calculated the contrast-to-noise ratio (CNR) as the difference between the mean uptake of gray and white matter, divided by the standard deviation of the difference, as follows:

$$\mathrm{CNR} = \frac{\bar{S}_g - \bar{S}_w}{\sqrt{\sigma_g^2 + \sigma_w^2}}, \tag{5}$$

where $\bar{S}_g$ and $\bar{S}_w$ are the mean activity of multiple frames in the corresponding the region of interest (ROI) of the gray and white matter and σ_g and σ_w are the standard deviations of multiple frames in the corresponding ROIs of the gray and white matter.

TABLE 1. Kinetic parameters used in the ^{18}F-FDG simulation.

Region	K1	k2	k3	k4
White matter	0.033	0.083	0.037	0
Gray matter	0.062	0.071	0.067	0

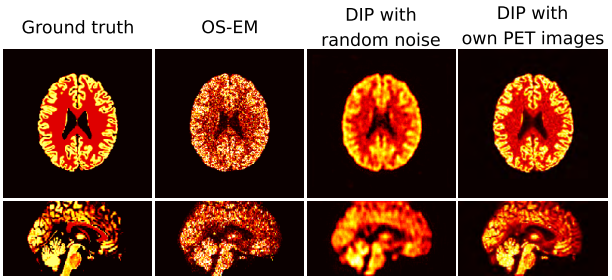

FIGURE 3. Simulation results of image reconstruction for frame 26. The columns correspond to the ground truth, reconstructed image using OS-EM algorithm, DIP method with random noise as the network input, and DIP method with the patient's own PET image (static PET image) as the network input (left to right).

V. RESULTS

A. COMPUTER SIMULATION

Figure 3 shows axial and sagittal images using random noise and a patient's own PET image as the network input for the simulation data for frame 26, correspond to 3900–4200 s, after the injection of ^{18}F-FDG. The image of this time frame mainly reflects a glucose metabolism. Using random noise as the network input, the dynamic PET image is smooth, however, the edges of the image, including the cortex structures, are lost. In contrast, when using the patient's own PET image as the network input, the PET image is smooth and the edges of the image remain visible. This result shows that the image quality is improved when using the patient's own data, compared to when random noise is used as the network input. We compared the DIP method with 3D GF, NLM, and the image-based dynamic IGF [17]. For GF, the full width at half maximum was set to 1.0 voxel. For NLM and IGF, the kernel size was set to $3 \times 3 \times 3$ voxels, respectively. Figure 4 shows the axial and sagittal images using different algorithms for the simulation data at frames 6 and 26, respectively. The images of frame 6 mainly reflect a cerebral blood flow. These time frames correspond to 120–160 s, and 3900–4200 s, respectively, after the injection of ^{18}F-FDG. We set a learning rate of 0.001 and 2500 epochs for the network training. The IGF and DIP methods incur less statistical noise and preserve the cortex structures compared with the GF and NLM algorithms. Especially, in the reconstructed images for an early phase, such as frame 6, the DIP method could better reduce the noise while preserving the structures, relative to the IGF algorithm. Figure 5 shows the regional TACs for the ROIs of the white and gray matter, while Table 2 lists the mean activities of each ROI for different methods correspond to

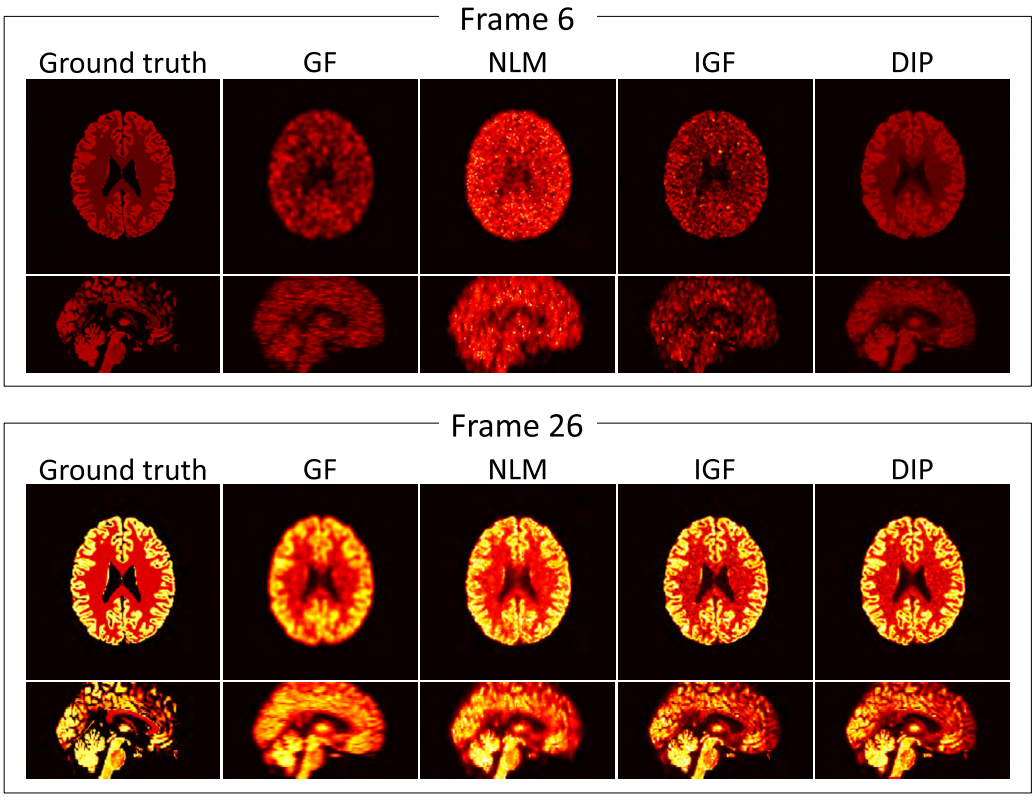

FIGURE 4. Simulation results of image reconstruction using OS-EM algorithm in frames 6 and 26. The columns correspond to the ground truth, GF, NLM, IGF, and DIP methods (left to right).

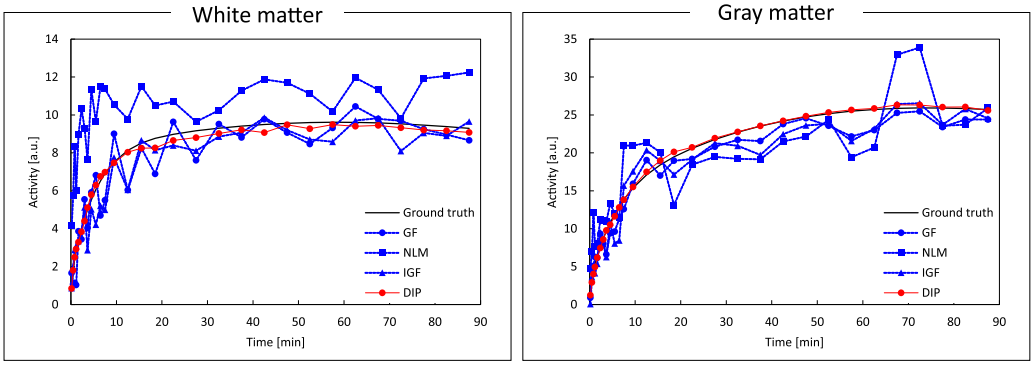

FIGURE 5. Simulation results for regional TACs determined using different methods for the corresponding voxels located in the white matter (left), and gray matter (right). Compared with the other algorithms, the regional TACs of the DIP method are closer to the ground truth.

TABLE 2. Mean activity of each ROI for different methods corresponding to 60–90 min, after injection of ^{18}F-FDG.

Method	Ground truth	GF	NLM	IGF	DIP
White matter [a.u.]	9.47	9.47 ± 0.64	11.56 ± 0.91	9.20 ± 0.66	9.27 ± 0.14
Gray matter [a.u.]	25.83	24.32 ± 0.99	26.79 ± 5.40	25.03 ± 1.44	26.03 ± 0.28

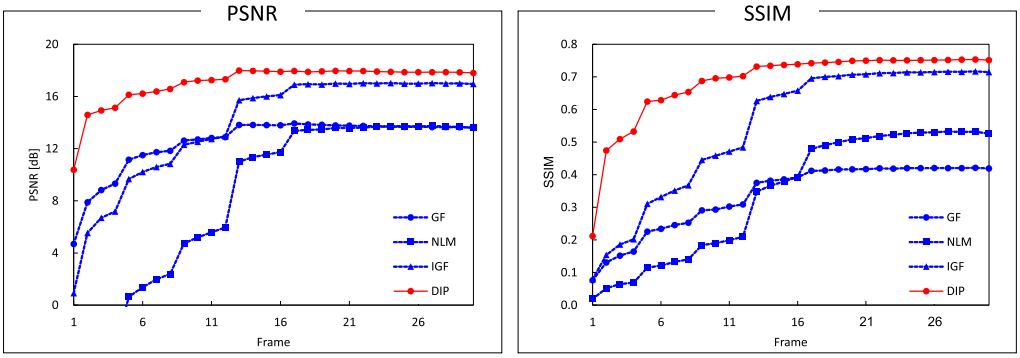

FIGURE 6. Performance evaluation metrics: PSNR (left), and SSIM (right) for each time frame, processed using different denoising algorithms. In the DIP method, the PSNR and SSIM were greater in all the time frames, relative to the other algorithms.

60–90 min, after the injection of ^{18}F-FDG. Compared with the other algorithms, the regional TACs of the DIP method are closer to the ground truth. Figure 6 shows the PSNR and SSIM curves for different methods corresponding to each time frame. With the DIP method, the PSNR and SSIM were greater in all the time frames, relative to the other algorithms. Figure 7 shows color maps of the PSNR as well as the SSIM of the number of epochs in each time frame. These results quantitatively show that the proposed DIP method can better reduce statistical noise, while preserving the edges, relative to the other algorithms.

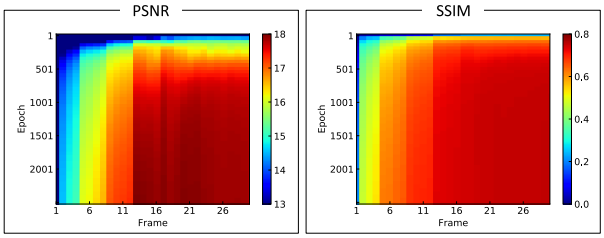

FIGURE 7. Color maps of the PSNR (left) and SSIM (right) of the number of epochs at each time frame.

B. PET MEASUREMENTS FOR LIVING MONKEY BRAIN

For GF, the full width at half maximum was set to 1.0 voxel. For both NLM and IGF, the kernel size was set to $3 \times 3 \times 3$ voxels. The MR image and the reconstructed PET images of the real dynamic ^{18}F-FDG PET data of the living monkey brain at frames 6, and 26, respectively, are shown in Figure 8. These time frames correspond to 120–160 s, and 3900–4200 s, respectively, after the injection of ^{18}F-FDG. We set a learning rate of 0.01 and 500 epochs for the network training. The DIP method was found to incur less noise and better preserve the cortex structures than the other algorithms. Figure 9 shows the regional TACs of the PET images treated by each algorithm. The regional TACs of the images treated using the DIP method were much smoother, relative to those processed with other algorithms. Table 3 shows the mean activity of each ROI and CNR for different methods, corresponding to 60–90 min after injection. The DIP method

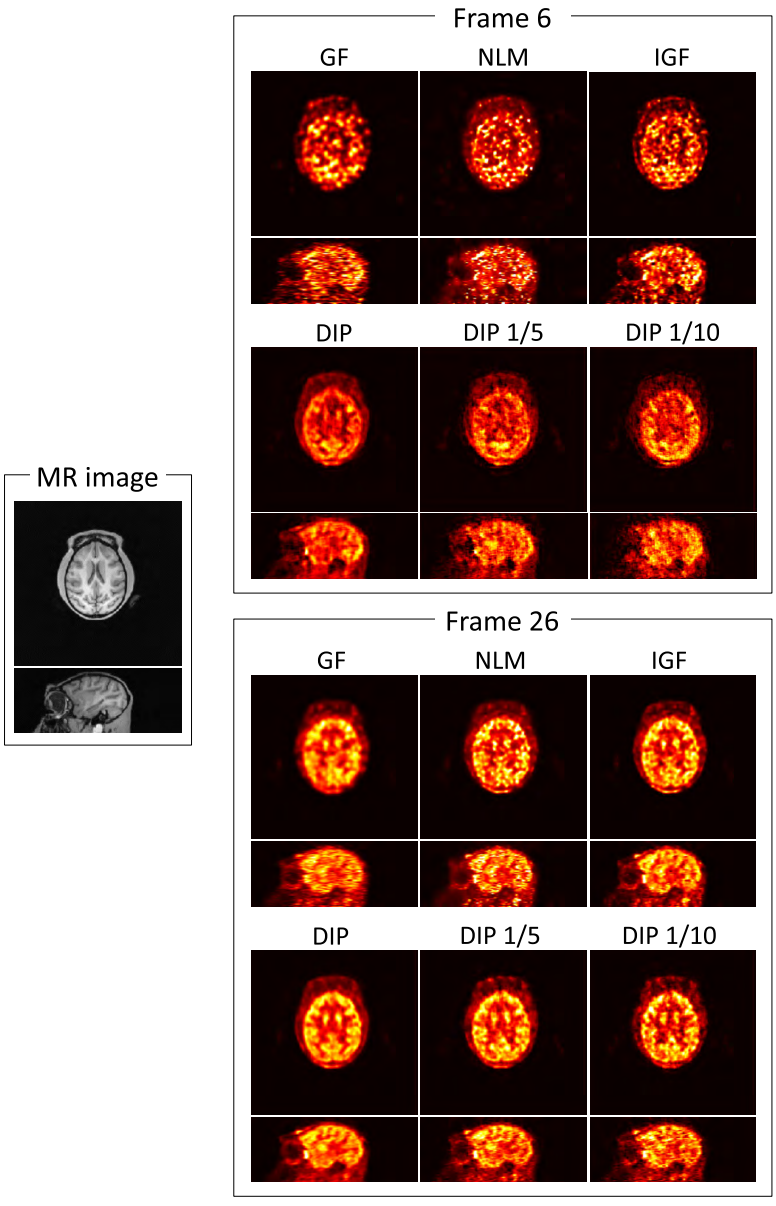

FIGURE 8. Reconstructed images of real dynamic 18F-FDG PET data of living monkey brain at frames 6, and 26.

produces a better CNR than the other algorithms, while the low-dose images treated by the DIP method also have better CNRs than those subjected to the other methods.

VI. DISCUSSION

Various deep-learning methods based on CNNs have been investigated for PET imaging [42], [43]. However, these methods are difficult to apply in a clinical setting if unknown cases, not included in the training datasets, are presented. Our method is capable of dealing with unknown cases because the original PET data itself is used to reduce the statistical noise.

In the computer simulation, the DIP method provided a higher PSNR than the other algorithms [17], [29], regardless of the time frame. The SSIM can evaluate the structural similarity, which cannot be evaluated by the PSNR [44]. Our DIP method also provides a higher SSIM than the other algorithms, regardless of the time frame. According to ref [22], CNNs have a knowledge about the structure of images. In fact, when compared to the IGF algorithm [17], the proposed method clearly depicts the cortex structures even in the early phases. In addition, the regional TACs of the DIP method are closer to the ground truth than with the other algorithms. These results suggest that the structure of the CNNs has an intrinsic ability to regularize the PET image denoising task compared with other ''hand craft'' type regularization [45].

One of the main limitations of dynamic PET is the resulting low-quality of the images, caused by statistical noise resulting

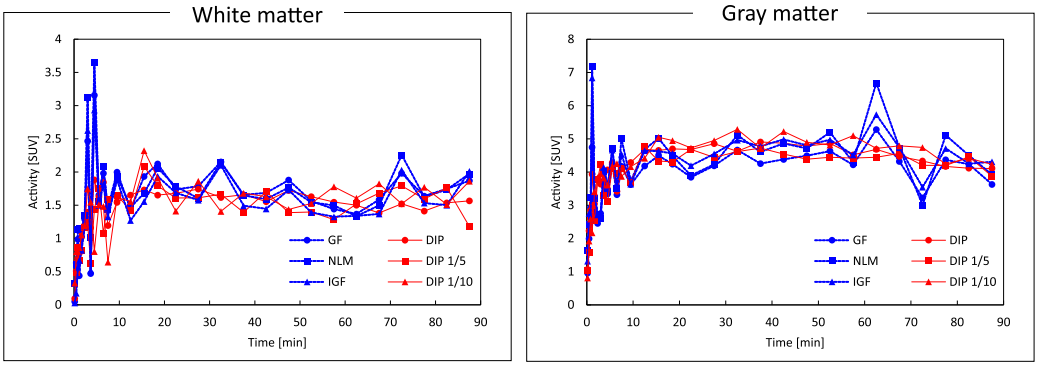

FIGURE 9. Regional ^{18}F-FDG TACs determined using different methods for the corresponding voxels located in the white (left) and gray matter (right) of a living monkey brain. The regional TACs of the images treated using the DIP method were much smoother, relative to those processed using other algorithms.

TABLE 3. Mean activity and contrast-to-noise ratio for different methods corresponding to 60–90 min, after injection of ^{18}F-FDG.

Method	GF	NLM	IGF	DIP		
Thinning rate	1	1	1	1	1/5	1/10
White matter [SUV]	1.70 ± 0.22	1.74 ± 0.33	1.63 ± 0.31	1.49 ± 0.07	1.59 ± 0.22	1.68 ± 0.16
Gray matter [SUV]	4.18 ± 0.71	4.66 ± 1.23	4.55 ± 0.72	4.31 ± 0.23	5.05 ± 0.23	4.91 ± 0.41
CNR	3.36	2.29	3.70	11.71	10.74	7.44

from the short time frames. In the real ^{18}F-FDG data, obtained from the living monkey brain, the DIP method subjectively reduced the noise and preserved the brain structure. Furthermore, the regional TACs of the white and gray matter in the DIP method are smoother than those produced by the other algorithms. Particularly, in the first 10 min after the ^{18}F-FDG injection, the variation resulting from the use of the DIP method is clearly smaller than that produced by the other algorithms. In the ideal metabolic trapping of ^{18}F-FDG kinetics, the image contrast between the gray and white matter is higher in the later time frames such from 60–90 min. As shown in the CNR (Table 3), in the DIP method, the contrast between the gray and white matter is clearly expressed relative to the other algorithms. In addition, the DIP method can maintain the CNR, even for low dose images, by sampling list data to 1/5 and 1/10 of original. These results suggest that the DIP method strongly improves the accuracy of the kinetics analysis and therefore would be useful for low-dose PET imaging.

This study was subject to several limitations. First, we chose and modified the 3D U-net architecture which has exhibited good performance in the field of biomedical applications [31]. The network structure had to be optimized to attain a high expression performance. Second, we evaluated only ^{18}F-FDG, for which the radioactivity distribution in the images does not change between the early and later time frames. Therefore, each dynamic PET image has a very similar distribution pattern. In the future, we will investigate the effect of the reversible-type PET tracers such as ^{11}C-raclopride. Finally, in the DIP method, the number of epochs was empirically determined. As shown in Figure 7, there is a plateau in the later epochs. We can detect this portion and automatically detect the number of epochs.

VII. CONCLUSION

In the present study, we examined a novel method for the post-reconstruction denoising of dynamic PET images using CNN without prior training datasets. Anatomical information such as MR or CT images was not necessary, as the original PET data itself was used to reduce the statistical noise. Computer simulation based on ^{18}F-FDG kinetics indicated that the DIP method reduced the statistical noise, while preserving the cortex structures, and achieved an improved quantitative TAC accuracy, compared to the other algorithms. In addition, real ^{18}F-FDG data, obtained from a living monkey brain, indicated that the DIP method outperformed the other algorithms in terms of CNR. Furthermore, the DIP method could maintain the CNR even for low-dose images by resampling the list data to 1/5 and 1/10 of the original, demonstrating that the DIP method could be applied to low-dose PET imaging. In the near future, we intend to evaluate the performance of the DIP method when using other PET tracers and human clinical PET data.

ACKNOWLEDGMENT

The authors would like to thank the scientific advice provided by Hideaki Tashima from National Institute of Radiological Sciences, National Institutes for Quantum and Radiological Science and Technology, Norihiro Harada and Takashi Isobe from the Central Research Laboratory, Hamamatsu photonics K. K. They would also like to thank the English proofing provided by Kosuke Kawamura of the Intellectual Property Division, Hamamatsu Photonics K. K.

REFERENCES

[1] M. E. Phelps, *PET: Molecular Imaging and Its Biological Applications*. New York, NY, USA: Springer, 2012.

[2] M. Watanabe, A. Saito, T. Isobe, K. Ote, R. Yamada, T. Moriya, and T. Omura, "Performance evaluation of a high-resolution brain PET scanner using four-layer MPPC DOI detectors," *Phys. Med. Biol.*, vol. 62, no. 17, pp. 7148–7166, 2017.

[3] H. Tashima, E. Yoshida, Y. Iwao, H. Wakizaka, T. Maeda, C. Seki, Y. Kimura, Y. Takado, M. Higuchi, T. Suhara, T. Yamashita, and T. Yamaya, "First prototyping of a dedicated PET system with the hemisphere detector arrangement," *Phys. Med. Biol.*, vol. 64, no. 6, Mar. 2019, Art. no. 065004.

[4] T. Okamoto, K. Ote, K. Sakai, A. Noda, K. Shimizu, K. Masuda, T. Ohmura, and M. Watanabe, "An animal PET scanner using flat-panel position-sensitive PMTs," *Ann. Nucl. Med.*, vol. 28, no. 1, pp. 74–80, Jan. 2014.

[5] P. J. Green, "Bayesian reconstructions from emission tomography data using a modified EM algorithm," *IEEE Trans. Med. Imag.*, vol. 9, no. 1, pp. 84–93, Mar. 1990.

[6] E. U. Mumcuoglu, R. M. Leahy, and S. R. Cherry, "Bayesian reconstruction of PET images: Methodology and performance analysis," *Phys. Med. Biol.*, vol. 41, no. 9, pp. 1777–1807, 1996.

[7] I.-T. Hsiao, A. Rangarajan, and G. Gindi, "A new convex edge-preserving median prior with applications to tomography," *IEEE Trans. Med. Imag.*, vol. 22, no. 5, pp. 580–585, May 2003.

[8] 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, 2012.

[9] C. Comtat, P. E. Kinahan, J. A. Fessler, T. Beyer, D. W. Townsend, M. Defrise, and C. Michel, "Clinically feasible reconstruction of 3D whole-body PET/CT data using blurred anatomical labels," *Phys. Med. Biol.*, vol. 47, no. 1, pp. 1–20, 2002.

[10] J. Tang and A. Rahmim, "Bayesian PET image reconstruction incorporating anato-functional joint entropy," *Phys. Med. Biol.*, vol. 54, no. 23, pp. 7063–7075, 2009.

[11] 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.

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

[13] 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, 2014, Art. no. e89282. doi: 10.1371/journal.pone.0089282.

[14] A. Buades, B. Coll, and J.-M. Morel, "A non-local algorithm for image denoising," in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, Jun. 2005, pp. 60–65. doi: 10.1109/CVPR.2005.38.

[15] J. Dutta, R. M. Leahy, and Q. Li, "Non-local means denoising of dynamic PET images," *PLoS ONE*, vol. 8, no. 12, 2013, Art. no. e81390. doi: 10.1371/journal.pone.0081390.

[16] 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, 2010.

[17] 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.

[18] M. Shidahara, Y. Ikoma, J. Kershaw, Y. Kimura, M. Naganawa, and H. Watabe, "PET kinetic analysis: Wavelet denoising of dynamic PET data with application to parametric imaging," *Ann. Nucl. Med.*, vol. 21, no. 7, p. 379, 2007.

[19] 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.

[20] K. Suzuki, "Overview of deep learning in medical imaging," *Radiol. Phys. Technol.*, vol. 10, no. 3, pp. 257–273, 2017.

[21] F. Hashimoto, A. Kakimoto, N. Ota, S. Ito, and S. Nishizawa, "Automated segmentation of 2D low-dose CT images of the psoas-major muscle using deep convolutional neural networks," *Radiol. Phys. Technol.*, vol. 12, no. 2, pp. 210–215, 2019.

[22] J. Cui, X. Liu, Y. Wang, and H. Liu, "Deep reconstruction model for dynamic PET images," *PLoS ONE*, vol. 12, no. 9, 2017, Art. no. e0184667.

[23] I. Häggström, C. R. Schmidtlein, G. Campanella, and T. J. Fuchs, "DeepPET: A deep encoder–decoder network for directly solving the PET image reconstruction inverse problem," *Med. Image Anal.*, vol. 54, pp. 253–262, May 2019.

[24] J. Lehtinen, J. Munkberg, J. Hasselgren, S. Laine, T. Karras, M. Aittala, and T. Aila, "Noise2Noise: Learning image restoration without clean data," Oct. 2018, *arXiv:1803.04189*. [Online]. Available: https://arxiv.org/abs/1803.04189

[25] D. Ulyanov, A. Vedaldi, and V. Lempitsky, "Deep image prior," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2018, pp. 9446–9454. doi: 10.1109/CVPR.2018.00984.

[26] G. Mataev, M. Elad, and P. Milanfar, "DeepRED: Deep image prior powered by RED," May 2019, *arXiv:1903.10176*. [Online]. Available: https://arxiv.org/abs/1903.10176

[27] 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. doi: 10.1109/TMI.2018.2888491.

[28] J. Cui, K. Gong, N. Guo, K. Kim, H. Liu, and Q. Li, "CT-guided PET parametric image reconstruction using deep neural network without prior training data" *Proc. SPIE*, vol. 10948, Mar. 2019, Art. no. 109480Z. doi: 10.1117/12.2513077.

[29] P. Coupé, P. Yger, S. Prima, P. Hellier, C. Kervrann, and C. Barillot, "An optimized blockwise nonlocal means denoising filter for 3-D magnetic resonance images," *IEEE Trans. Med. Imag.*, vol. 27, no. 4, pp. 425–441, Apr. 2008.

[30] E. Garyfallidis, M. Brett, B. Amirbekian, A. Rokem, S. van der Walt, M. Descoteaux, I. Nimmo-Smith, and D. Contributors, "Dipy, a library for the analysis of diffusion MRI data," *Frontiers Neuroinform.*, vol. 8, p. 8, Apr. 2014. doi: 10.3389/fninf.2014.00008.

[31] Ö. Çiçek, A. Abdulkadir, S. S. Lienkamp, T. Brox, and O. Ronneberger, "3D U-Net: Learning dense volumetric segmentation from sparse annotation," in *Medical Image Computing and Computer-Assisted Intervention (MICCAI)* (Lecture Notes in Computer Science). Cham, Switzerland: Springer, 2016.

[32] K. He, X. Zhang, S. Ren, and J. Sun, "Delving deep into rectifiers: Surpassing human-level performance on imagenet classification," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Dec. 2015, pp. 1026–1034. doi: 10.1109/ICCV.2015.123.

[33] M. Abadi. *TensorFlow: A System for Large-Scale Machine Learning*. Accessed: Mar. 2019. [Online]. Available: https://www.usenix.org/system/files/conference/osdi16/osdi16-abadi.pdf

[34] *Keras: The Python Deep Learning Library*. Accessed: Mar. 2019. [Online]. Available: https://keras.io/

[35] B. Aubert-Broche, M. Griffin, G. P. 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.

[36] D. Feng, K.-P. Wong, C.-M. Wu, and W.-C. Siu, "A technique for extracting physiological parameters and the required input function simultaneously from PET image measurements: Theory and simulation study," *IEEE Trans. Inf. Technol. Biomed.*, vol. 1, no. 4, pp. 243–254, Dec. 1997.

[37] M. C. Huisman, L. W. van Golen, N. J. Hoetjes, H. N. Greuter, P. Schober, R. G. Ijzerman, M. Diamant, and A. A. Lammertsma, "Cerebral blood flow and glucose metabolism in healthy volunteers measured using a high-resolution PET scanner," *EJNMMI Res.*, vol. 2, no. 1, p. 63, Nov. 2012.

[38] H. M. Hudson and R. S. Larkin, "Accelerated image reconstruction using ordered subsets of projection data," *IEEE Trans. Med. Imag.*, vol. 13, no. 4, pp. 601–609, Dec. 1994.

[39] S. van der Walt, J. L. Schönberger, J. Nunez-Iglesias, F. Boulogne, J. D. Warner, N. Yager, E. Gouillart, and T. Yu, "Scikit-image: Image processing in Python," *PeerJ*, vol. 2, p. e453, Jun. 2014.

[40] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, "Image quality assessment: From error visibility to structural similarity," *IEEE Trans. Image Process.*, vol. 13, no. 4, pp. 600–612, Apr. 2004.

[41] E. Tanaka and H. Kudo, "Optimal relaxation parameters of DRAMA (dynamic RAMLA) aiming at one-pass image reconstruction for 3D-PET," *Phys. Med. Biol.*, vol. 55, no. 10, pp. 2917–2939, 2010.

[42] 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.

[43] 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.

[44] Z. Wang and A. C. Bovik, "Mean squared error: Love it or leave it? A new look at signal fidelity measures," *IEEE Signal Process. Mag.*, vol. 26, no. 1, pp. 98–117, Jan. 2009.
[45] L. I. Rudin, S. Osher, and E. Fatemi, "Nonlinear total variation based noise removal algorithms," *Phys. D, Nonlinear Phenomena*, vol. 60, nos. 1–4, pp. 259–268, 1992.

FUMIO HASHIMOTO received the B.S. and M.S. degrees in health science from Fujita Health University, Toyoake, Japan, in 2016. He is currently a Researcher with the Central Research Laboratory, PET Research Group, Hamamatsu Photonics K.K. His research interests include medical image reconstruction and processing.

HIROYUKI OHBA received the Ph.D. degree in engineering from Shizuoka University, Japan, in 1996. He is currently a Senior Researcher with the Central Research Laboratory, PET Research Group, Hamamatsu Photonics K.K. His research interests include the kinetic analysis of PET tracer in living brain and PET image processing.

KIBO OTE received the B.S. and M.S. degrees from Shizuoka University, Hamamatsu, Japan, in 2000. He is currently an Associate Senior Researcher with the Central Research Laboratory, 5th Research Group, Hamamatsu Photonics K.K. His research interests include medical image reconstruction and processing.

ATSUSHI TERAMOTO received the B.S. and M.S. degrees in electric engineering from Meijo University, Nagoya, Japan, in 1998, and the Ph.D. degree in engineering from Gifu University, in 2008. From 1998 to 2008, he was a Research Engineer with Nagoya Electric Works Company Ltd. From 2008 to 2012, he was an Assistant Professor with Fujita Health University, where he is currently an Associate Professor. His research interests include artificial intelligence and image processing.

HIDEO TSUKADA received the B.S., M.S., and Ph.D. degrees from the Department of Pharmacology, Shizuoka Collage of Pharmacy, Shizuoka, Japan, in 1982, 1984, and 1987, respectively. He is currently with the Central Research Laboratory, Hamamatsu Photonics K.K. He was sent to the National Institute of Radiological Science, Chiba, Japan, on a mission to research positron emission tomography (PET), from 1988 to 1990. From 1990 to 1991, he was sent to the PET Centre, Uppsala University, directed by Prof. B. Langstrom, Uppsala, Sweden, to research the applications of PET technology for drug discovery and development. In 1991, he returned to Hamamatsu, and established the PET Center, Hamamatsu Photonics, in 1992, which was the first PET facility owned by a private company, and being dedicated for basic PET research with experimental animals not for clinical use and conducted PET studies from preclinical to clinical stages. He has been a Visiting Professor with the Hamamatsu University School of Medicine, Hamamatsu, since 2007, and the School of Pharmaceutical Sciences, University of Shizuoka, Shizuoka, since 2006. He is currently an Adviser of CREST and ERATO projects of Japan Science and Technology Agency, Tokyo, Japan. He has published over 250 papers. He has received the Society for Nuclear Medicine, in 2009, and the Japan Molecular Imaging Award, in 2010.