Dynamic PET Image Reconstruction Using Nonnegative Matrix Factorization Incorporated With Deep Image Prior

Tatsuya Yokota[*,⋆], Kazuya Kawai[*], Muneyuki Sakata[†], Yuichi Kimura[‡], and Hidekata Hontani[*]
[*] Nagoya Institute of Technology, Nagoya, Japan, {t.yokota, hontani}@nitech.ac.jp
[⋆] RIKEN Center for Advanced Intelligence Project, Tokyo, Japan
[†] Tokyo Metropolitan Institute of Gerontology, Tokyo, Japan
[‡] Kindai University, Wakayama, Japan

Abstract

We propose a method that reconstructs dynamic positron emission tomography (PET) images from given sinograms by using non-negative matrix factorization incorporated with a deep image prior for appropriately constraining the spatial patterns of resultant images. The proposed method can reconstruct dynamic PET images with higher signal-to-noise ratio and blindly decompose an image matrix into pairs of spatial and temporal factors. The former represent homogeneous tissues with different kinetic parameters and the latter represent the time activity curves that are observed in the corresponding homogeneous tissues. We employ U-Nets combined in parallel for deep image prior and each of the U-Nets is used to extract each spatial factor decomposed from the data matrix. Experimental results show that the proposed method outperforms conventional methods and can extract spatial factors that represent the homogeneous tissues.

1. Introduction

Positron emission tomography (PET) is the most modern nuclear medicine modality [27] and is used for observing biomolecular mechanisms such as the metabolic processes in the body. The biomolecular mechanisms at each location in the body can be represented by a set of physiological kinetic parameters and their parameter values can be estimated from the temporal change of each voxel value in dynamic PET images, which is a temporal series of 3D static PET images [20, 19, 32, 10, 25, 9]; Each 3D static PET image represents the spatial distribution of the concentration of radioactive tracer, and the temporal change of each voxel value, called as a time-activity curve (TAC), of dynamic PET describes the kinetics of the radioactive tracer at each location in the body. The accuracy of the estimation of the kinetic parameters largely changes depending on the signal-to-noise ratio (SNR) of a TAC of each voxel in given dynamic PET images, and it is still a challenging problem to reconstruct high-quality dynamic PET images.

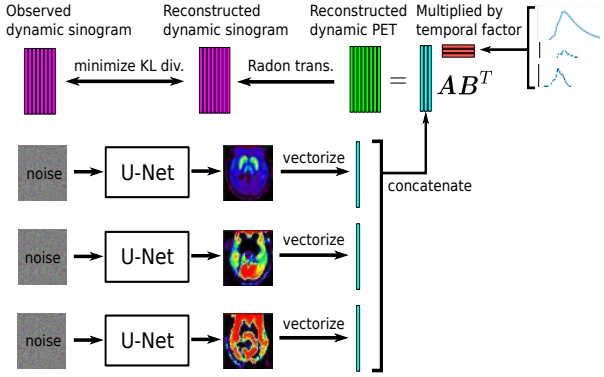


Figure 1. Overview of the proposed PET image reconstruction method with U-Nets for the deep image prior (DIP)

A set of dynamic PET images is reconstructed from a temporal series of sinograms, each of which is a Radon transformation of the spatial concentration distribution of the radio active tracer and is measured by a PET scanner at each time. It is known that the PET image reconstruction is highly sensitive to measurement noises and is an ill-posed problem [28]. Models that appropriately constrain PET images are therefore needed for stable reconstruction of dynamic PET images.

In this study, we discuss the priors/models for dynamic PET images from spatial, temporal, and spatio-temporal aspects. First, the *spatio-temporal low-rank structure* of dynamic PET images should be considered. This means that the matrix representation of dynamic PET image, $\boldsymbol{Z} \in \mathbb{R}_{+}^{N_{\mathrm{i}} \times T}$, is low-rank, where N_{i} denotes the number of the voxels and T denotes the number of frames. In the human body, there exist homogeneous tissues; in each of them the temporal tracer behavior is the same and the TACs are the

same at any location [6, 27]. The variety of TACs observed in a dynamic PET image should therefore be much less than the number of voxels, N_{i}. In addition, the tracer activity can be well represented by a system of ordinary differential equations, a compartment model [32], that has fewer kinetic parameters (usually 2 or 4) than the number of frames, T.

For exploiting the low-rank structure of dynamic PET image, we consider non-negative matrix factorization (NMF) [12, 11], $\boldsymbol{Z} = \boldsymbol{A}\boldsymbol{B}^T$, since the concentration of radioactive tracer is absolutely *non-negative*, and individual homogeneous tissues in the body should be *spatially independent*. These non-negativity and spatial independency of homogeneous tissues can be modeled by *exclusive sparseness* of the non-negative factor matrix $\boldsymbol{A} = [\boldsymbol{a}_1, ..., \boldsymbol{a}_R]$.

Furthermore, we introduce the spatial model of $\boldsymbol{a}_r (r = 1, ..., R)$ for constraining the spatial patterns of images to natural ones. One of the most major constraints is realized by the total variation (TV)-norm regularization; this is widely used in various ill-posed problems such as super-resolution [29], inpainting [8], blind deconvolution [4], and also PET image reconstruction [33, 2]. Our proposed method, though, does not employ these conventional models for the representation of the spatial aspects but does employ *deep image prior* (DIP) [30], which have very strong ability of representing natural images. One advantage of DIP is that, even though deep neural networks (DNNs) are used, no training data is required and hence DIPs can avoid potential biases arising from training-based regularizations; it is not easy to collect numerous medical images for constructing image priors that have sufficient ability to represent generic patterns of medical images. In [7], DIP was firstly applied for static PET image reconstruction. By contrast, we use DIP for 'spatial bases' in NMF based dynamic PET image reconstruction.

As for the temporal factor matrix $\boldsymbol{B}$, the proposed method introduces a model that prefers smooth TACs [24]. Many methods for dynamic PET image reconstruction directly reconstruct all frames from all sinograms observed at different time frames so that the TACs are smooth at all voxels. TACs can be made smooth by iteratively smoothing TACs during the reconstruction process [31] or by representing TACs with smooth basis functions [21, 22]. The quadratic variation (QV) of TACs can also be used for the constraint of the solution space [11]. The proposed method employs the QV of TACs as a model for obtaining smooth TACs.

In summary, we propose a method that reconstructs dynamic PET images with high SNR by using NMF incorporated with DIP. The proposed method can also spatially separate the body region into homogeneous tissues and obtain the noise reduced TAC for each of the tissues. The separated tissue regions and corresponding TACs are represented by the blindly decomposed factors computed by the NMF, which can generate sparse factors [16, 17]. The DIP is used to make the spatial factors natural. The smoothness of TACs, the DIP, and the low-rank of image matrices are all integrated into one optimization problem for the dynamic PET image reconstruction. The contributions of this study are as follows:

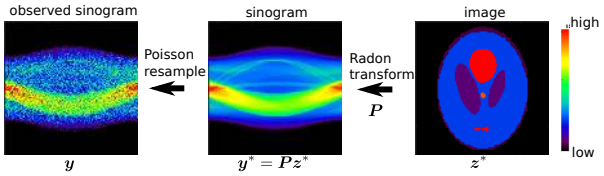


Figure 2. Sinogram and image, where $\boldsymbol{y} = \boldsymbol{P}\boldsymbol{z}^* + \boldsymbol{n}$.

1. The DIP is incorporated into NMF-based dynamic PET image reconstruction as shown in Fig.1 and the optimization algorithm for PET image reconstruction is explicitly derived.
2. The reconstruction accuracy is quantitatively compared between the proposed method and the other existing methods using artificial images and clinical images of the brain. This demonstrates that the DIP is also useful for PET image reconstruction.
3. As far as we know, this is the first study to successfully separate brain regions, and basis of TACs from only sinograms without using other auxiliary information like CT/MRI.

2. PET image reconstruction

2.1. Static PET image reconstruction

The sinogram $\boldsymbol{y}$ is observed by the following system:

$$\boldsymbol{y} = \boldsymbol{P}\boldsymbol{z}^* + \boldsymbol{n}, \tag{1}$$

where $\boldsymbol{y} \in \mathbb{R}_+^{N_{\mathrm{s}}}$ is a vector representation of a 2D/3D-sinogram, $\boldsymbol{P} \in \mathbb{R}_+^{N_{\mathrm{s}} \times N_{\mathrm{i}}}$ is a linear operator for performing the Radon transform, $\boldsymbol{z}^* \in \mathbb{R}_+^{N_{\mathrm{i}}}$ is a vector representation of 2D/3D true spatial distribution of radioactivity, and $\boldsymbol{n} \in \mathbb{R}^{N_{\mathrm{s}}}$ is a noise factor that satisfies $\boldsymbol{y} \sim \mathrm{Poisson}(\boldsymbol{y}^*)$, where $\boldsymbol{y}^* = \boldsymbol{P}\boldsymbol{z}^*$. N_s and N_i are respectively number of voxels of sinogram and PET image. Figure 2 shows the relationship between the sinogram and PET image. PET image reconstruction estimates the PET image, $\boldsymbol{z}^*$, from the measured sinogram, $\boldsymbol{y}$, given that the Radon transformation, $\boldsymbol{P}$, is known in advance.

There are a few fundamental methods for PET image reconstruction. Filtered back projection (FBP) is the simplest method [1, 23] and is widely used. It can be written as

$$\widehat{\boldsymbol{z}}_{\mathrm{FBP}} := \boldsymbol{P}^T \boldsymbol{F} \boldsymbol{y}, \tag{2}$$

where $\boldsymbol{F}$ is a linear operator to perform some filtering process such as ramp, Hamming, and Hanning filters. It is well known that the FBP often suffers from measurement noise.

The second fundamental method is maximum likelihood estimation [15, 28] which is described by

$$\widehat{\boldsymbol{z}}_{\mathrm{EM}} := \underset{\boldsymbol{z}\geq 0}{\operatorname{argmin}}\, D_{\mathrm{KL}}(\boldsymbol{y}||\boldsymbol{P}\boldsymbol{z}), \tag{3}$$

where $D_{\mathrm{KL}}(\boldsymbol{p}||\boldsymbol{q}) := \sum_i p_i \log(p_i/q_i)$ is a Kullback-Leibler (KL) divergence. This can be solved by the expectation-maximization (EM) algorithm. For example, $\widehat{\boldsymbol{z}}_{\mathrm{EM}}$ can be strictly obtained by the multiplicative update system: $\boldsymbol{z} \leftarrow \boldsymbol{z} \circledast \{\boldsymbol{P}^T(\boldsymbol{y} \oslash \boldsymbol{P}\boldsymbol{z})\} \oslash (\boldsymbol{P}^T\mathbf{1}_{N_{\mathrm{i}}})$, where $\circledast$ and $\oslash$ are respectively the element-wise product and division operators, and $\mathbf{1}_{N_{\mathrm{i}}}$ is a N_{i}-dimensional vector whose elements are all one. Although the sinogram reconstructed by the EM algorithm can extremely fit the observed sinogram, the reconstructed PET image is highly sensitive to measurement noise and hence small-scale smoothing is often applied to the resultant images to improve their SNR.

For preventing the noise effect of PET image reconstruction, an *image prior* could be useful. As a good example, TV regularization can help to reconstruct the smoother images:

$$\widehat{\boldsymbol{z}}_{\mathrm{TV}} := \underset{\boldsymbol{z}\geq 0}{\operatorname{argmin}}\, D_{\mathrm{KL}}(\boldsymbol{y}||\boldsymbol{P}\boldsymbol{z}) + \lambda||\mathrm{img}(\boldsymbol{z})||_{\mathrm{TV}}, \tag{4}$$

where $\mathrm{img}(\boldsymbol{z})$ is a 2D/3D image (matrix/tensor) representation of a PET image, $||\cdot||_{\mathrm{TV}}$ is a TV-norm which stands for the non-smoothness of an image, and λ is a trade-off parameter for balancing the sinogram consistency and image consistency. Problem (4) is a convex optimization problem and recent convex optimization frameworks such as proximal splitting methods [3, 5] can solve it efficiently. A bottleneck of the TV prior is removing the detailed texture of an image when the trade-off parameter λ is too large or measurement noise is heavy.

2.2. Dynamic PET image reconstruction

A representation of a dynamic PET image contains an additional axis for the temporal domain. The system of observing the sinogram is rewritten as

$$\boldsymbol{Y} = \boldsymbol{P}\boldsymbol{Z}^* + \boldsymbol{N}, \tag{5}$$

where $\boldsymbol{Y} = [\boldsymbol{y}_1, ..., \boldsymbol{y}_T] \in \mathbb{R}_+^{N_{\mathrm{s}}\times T}$ and $\boldsymbol{Z}^* = [\boldsymbol{z}_1^*, ..., \boldsymbol{z}_T^*] \in \mathbb{R}_+^{N_{\mathrm{i}}\times T}$ are matrix representations of the dynamic sinogram and PET images, and $\boldsymbol{N}$ is a Poisson noise matrix. The merit of dynamic PET is the existence of common components (spatial and temporal patterns) which can be utilized for robust reconstruction. In [12, 11], the authors proposed a method to reconstruct the dynamic PET image matrix by NMF: $\boldsymbol{Z} = \boldsymbol{A}\boldsymbol{B}^T$, where $\boldsymbol{A} = [\boldsymbol{a}_1, ..., \boldsymbol{a}_R] \in \mathbb{R}_+^{N_{\mathrm{i}}\times R}$ and $\boldsymbol{B} = [\boldsymbol{b}_1, ..., \boldsymbol{b}_R] \in \mathbb{R}_+^{T\times R}$ are respectively the spatial and temporal factor matrices. Here, $R \ll N_{\mathrm{i}}$ and $R \ll T$ are satisfied.

The NMF would generate sparse spatial factors [16, 17], $\boldsymbol{a}_r$, and each corresponding temporal factor, $\boldsymbol{b}_r$, represents the temporal activity curve (TAC) observed in the tissues represented by the nonzero components in $\boldsymbol{a}_r$. The method proposed in [12] constrains the spatial patterns of PET images to linear combinations of low-frequency discrete cosine transform bases. This constraint can suppress noises in reconstructed images and improve their SNR but blurs images and often generates ringing artifacts. The reconstruction performance can be improved by using another prior that has better representation ability.

We therefore incorporate the DIP into the NMF framework as the spatial pattern prior.

3. Proposed method

3.1. Preliminary study of DIP for PET images

The DIP can be used in the image restoration problem by utilizing properties of convolutional neural network (CNN) architecture. In [30], the authors claimed that the CNN architecture used for the image generator contains a prior of an image. When a random noise is input to a randomly initialized CNN and the network coefficients are updated so that the CNN outputs a target *corrupted* image then, surprisingly, the CNN automatically restores the *uncorrupted* image before the output image completely fits the corrupted one. In other words, the CNN architecture itself is a prior.

The direct application of DIP for PET image reconstruction can be written by

$$\underset{\theta}{\text{minimize}}\, D_{\mathrm{KL}}(\boldsymbol{y}||\boldsymbol{P}\boldsymbol{\phi}(u|\theta)), \tag{6}$$

where $\boldsymbol{\phi}(\cdot) \in \mathbb{R}_+^{N_{\mathrm{i}}}$ stands for the output of an image-generative CNN such as a U-Net, u is an input random noise for the CNN, and θ is a set of network parameters such as convolutional kernels and biases. Setting a sinogram, $\boldsymbol{y}$, as a target, our method updates the coefficients of CNN, which are initialized randomly, so that the network outputs a PET image, $\phi(u|\theta)$, that minimizes $D_{\mathrm{KL}}(\boldsymbol{y}\|\boldsymbol{P}\phi(u|\theta))$. As described in [30], we need to handcraft a network that has sufficient ability to represent target images. We designed the network architecture by setting a noise-free sinogram as a target and evaluating the SNR of the images outputted by each of the examined networks. Figure 3 shows some examples. We found a single chain of standard building blocks of ConvNets, i.e., convolutional layers with optional downsampling and skip connections, would output *edge-enhanced* images; hence, the SNR of the outputs was low. A simple strategy for improving the representation ability of CNN is to use multiple chains of building blocks rather than

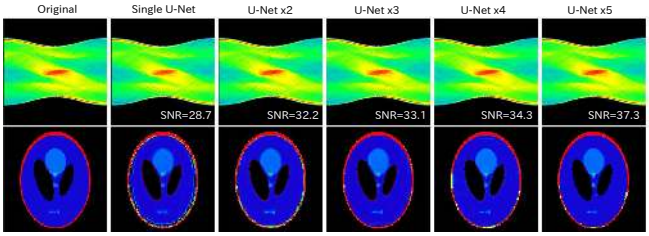


Figure 3. Simple application of deep image prior to PET image reconstruction. The leftmost column shows the original data. The obtained results in the case where the number of CNNs are 1, 2, 3, 4, and 5 are shown in order from the second column to the right.

designing a single but large and complex architecture. We combine multiple simple U-Nets [26] in parallel and solve the following problem for the reconstruction:

$$\begin{aligned} &\underset{\theta_1,\theta_2,\ldots,\theta_R}{\text{minimize}} \quad D_{\mathrm{KL}}(\boldsymbol{y}||\boldsymbol{P}\boldsymbol{z}), \\ &\text{subject to} \quad \boldsymbol{z} = \boldsymbol{\phi}(u|\theta_1) + \boldsymbol{\phi}(u|\theta_2) + \cdots + \boldsymbol{\phi}(u|\theta_R), \end{aligned} \tag{7}$$

where R is the number of U-Nets. Clearly, Problem (7) with $R = 1$ is equivalent to Problem (6). Figure 3 shows the results of PET image reconstruction from a noisefree sinogram by using multiple U-Nets with $R \in \{1, 2, ..., 5\}$. We can see the improvements of the representation ability of parallel U-Nets with larger R.

3.2. NMF incorporated with DIP

The proposed method reconstructs dynamic PET images by solving the following optimization problem. In our proposed model, each spatial factor of PET images, $\boldsymbol{a}_r$, is generated by each of the U-Nets, $\phi(u|\theta_r)$:

$$\begin{aligned} \underset{\Theta,\boldsymbol{B}}{\text{minimize}} \quad & \mathcal{L} := D_{\mathrm{KL}}(\boldsymbol{Y}||\boldsymbol{P}\boldsymbol{A}\boldsymbol{B}^T) \\ & \qquad + \alpha||\boldsymbol{A}^T||^2_{p,2} + \beta||\boldsymbol{B}||^2_{\mathrm{QV}}, \\ \text{s.t.} \quad & \boldsymbol{A} = [\boldsymbol{a}_1, ..., \boldsymbol{a}_R] \geq 0, \boldsymbol{B} \geq 0, \\ & \boldsymbol{a}_r = \boldsymbol{\phi}(u|\theta_r) \in [0, 1]^{N_{\mathrm{i}}}, \\ & ||\boldsymbol{a}_r||_\infty = 1 \text{ for } r = 1, 2, ..., R, \end{aligned} \tag{8}$$

where $\Theta = \{\theta_1, ..., \theta_R\}$ is a set of network parameters of CNNs, $|||[\boldsymbol{x}_1, ..., \boldsymbol{x}_N]|||^2_{p,2} := \sum_{n=1}^N ||\boldsymbol{x}_n||^2_p$ is a squared $l_{p,2}$-norm for the exclusiveness of spatial patterns, $||\boldsymbol{B}||_{\mathrm{QV}} := ||\boldsymbol{\nabla}\boldsymbol{B}||^2_F$ is a quadratic variation (QV) for smoothness of temporal patterns, and $\alpha \geq 0$ and $\beta \geq 0$ are weighting parameters of penalty terms. Let us use the differential operator $\boldsymbol{\nabla} = \boldsymbol{L}$ as

$$\boldsymbol{L} := \begin{pmatrix} 1 & -1 & 0 & \cdots & 0 \\ 0 & 1 & -1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \ddots & \vdots \\ 0 & \cdots & \cdots & 1 & -1 \end{pmatrix} \in \mathbb{R}^{(T-1)\times T}, \tag{9}$$

Table 1. SNR of sinogram reconstruction

	noisefree	40 dB	30 dB	20 dB	10 dB
FBP	36.57	36.31	27.90	18.38	7.26
EM	**69.53**	**44.88**	26.52	19.35	10.45
TV	44.58	43.60	35.98	29.51	13.16
Proposed	40.41	40.41	**39.31**	**34.76**	**15.58**

then we have $||\boldsymbol{B}||_{\mathrm{QV}} = \mathrm{trace}(\boldsymbol{B}^T\boldsymbol{L}^T\boldsymbol{L}\boldsymbol{B})$. The range of each $\boldsymbol{\phi}(u|\theta_r)$ is controlled by a sigmoid layer before output in the CNN. There are three hyper-parameters p, α, and β. Typically, we used $p = 0.5$, $\alpha = 0.01$, and $\beta \in \{0.001, 0.01, 0.1, 1.0, 10.0\}$ in our experiments. Figure 1 shows an overview of the proposed method in case $R = 3$.

For optimizing Problem (8), we employ the alternating optimization strategy in accordance with the NMF's research field. In other words, we alternately update Θ and $\boldsymbol{B}$. Thanks to the recent development of neural network optimization, Θ can be easily updated using the sophisticated gradient descent method with some library such as `TensorFlow`, `Caffe`, `PyTorch`, etc. We use the Adam optimizer implemented in `TensorFlow` with learning rate μ_Θ. Typically, we set $\mu_\Theta = 0.01$, and we decay $\mu_\Theta \leftarrow 0.98\mu_\Theta$ every 100 iterations.

On the other hand, $\boldsymbol{B}$ can be optimized by an EM-based multiplicative update rule which is usually used on many variants of NMF models.

$$\boldsymbol{B} \leftarrow \boldsymbol{B} \circledast [\partial\boldsymbol{B}_- \oslash \partial\boldsymbol{B}_+]^{\cdot\mu_{\boldsymbol{B}}}, \tag{10}$$

where $[\cdot]^{\cdot\mu}$ stands for the element-wise power with μ, and

$$\partial\boldsymbol{B}_- = (\boldsymbol{Y} \oslash \boldsymbol{P}\boldsymbol{A}\boldsymbol{B}^T)^T\boldsymbol{P}\boldsymbol{A} + \beta\max(-\boldsymbol{H}\boldsymbol{B}, 0), \tag{11}$$

$$\partial\boldsymbol{B}_+ = \boldsymbol{1}_{T\times N_{\mathrm{s}}}\boldsymbol{P}\boldsymbol{A} + \beta\max(\boldsymbol{H}\boldsymbol{B}, 0), \tag{12}$$

where $\boldsymbol{H} := \boldsymbol{L}^T\boldsymbol{L}$. Typically, we set $\mu_{\boldsymbol{B}} = 0.01$.

The optimization algorithm is summarized in Algorithm 1. According to the DIP strategy, we introduce the maximum value of iterations $I_{\max}$ for preventing the CNN from overfitting noisy targets. For robust CNN learning, we add a small noise v to the input of the CNNs at the ninth line in Algorithm 1. The eleventh line in Algorithm 1 is important for normalizing each spatial basis for improving the uniqueness of solutions and for preventing the divergence of the optimization. At the 13th line, $i_{\boldsymbol{B}}$ is the number of inner-iterations to update $\boldsymbol{B}$ for balancing both optimizations of Θ and $\boldsymbol{B}$. Since we set the step size of $\mu_{\boldsymbol{B}}$ to be relatively smaller for preventing the divergence of the algorithm, we set $i_{\boldsymbol{B}}$ relatively larger (typically, 10–100).

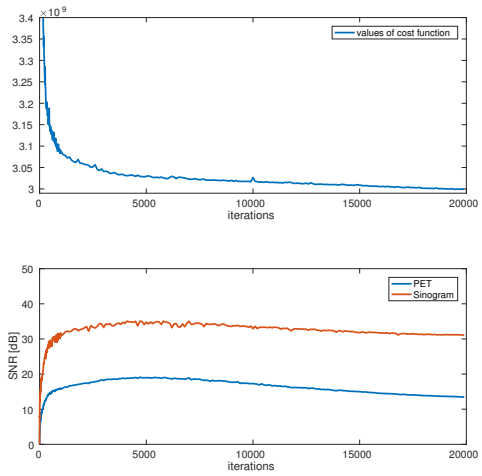
4. Experiments using simulation data

4.1. Simulation data

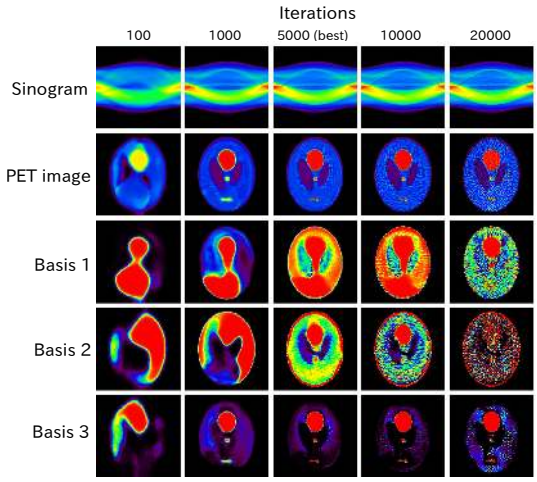
In this section, we show the experimental results of the proposed method by using simulation data for quantita-

Algorithm 1 Optimization algorithm

input: $\boldsymbol{Y}$, R, (p, α, β), $I_{\max}$, C, μ_Θ, $\mu_{\boldsymbol{B}}$, $i_{\boldsymbol{B}}$;
initialize:
  $k = 1$;
  $u \in [0, 0.1]^{N_\mathrm{i} \times C}$ is generated by random uniform distributions; (C is code depth)
  Θ^k is initialized randomly;
  $\boldsymbol{B}^k \leftarrow \boldsymbol{Y}^T \mathbf{1}_{N_\mathrm{s} \times R} + \boldsymbol{U}$, where $\boldsymbol{U} \sim$ Uniform(0,1); for all $r \in \{1, ..., R\}$;
repeat
  Generate random noise $v \in [0, 1/30]^{N_\mathrm{i} \times C}$ by random uniform distributions;
  Update $\Theta^{k+1} = [\theta_1^{k+1}, ..., \theta_R^{k+1}]$ for minimizing $\mathcal{L}(u + v|\Theta^k)$ with learning rate μ_Θ;
  $[\boldsymbol{a}_1^{k+1}, ..., \boldsymbol{a}_R^{k+1}] \leftarrow [\boldsymbol{\phi}(u|\theta_1^{k+1}), ..., \boldsymbol{\phi}(u|\theta_R^{k+1})]$;
  $\boldsymbol{a}_r^{k+1} \leftarrow \frac{\boldsymbol{a}_r^{k+1}}{\max(\boldsymbol{a}_r^{k+1})}$; for all $r \in \{1, ..., R\}$;
  $\boldsymbol{A}^{k+1} \leftarrow [\boldsymbol{a}_1^{k+1}, ..., \boldsymbol{a}_R^{k+1}]$;
  for $i_{\boldsymbol{B}}$ times **do**
    Update $\boldsymbol{B}^{k+1}$ by update rule (10) with step size $\mu_{\boldsymbol{B}}$;
  end for
  $k = k + 1$;
until $k > I_{\max}$
output: $\boldsymbol{A}^k$, $\boldsymbol{B}^k$



(a) Cost and reconstruction accuracy



(b) Sinogram, PET image, and spatial factors, $\boldsymbol{a}_r^k$

Figure 4. Example of the results obtained during the optimization process with inputting a noisy sinogram (SNR = 20dB)

Table 2. SNR of PET image reconstruction

	noisefree	40 dB	30 dB	20 dB	10 dB
FBP	19.53	19.23	10.41	0.85	-9.24
EM	**29.43**	18.96	3.89	-1.44	-6.99
TV	22.22	22.20	18.22	14.84	3.07
Proposed	26.44	**26.60**	**24.62**	**18.96**	**6.21**

Table 3. Selected values of parameters

	noisefree	40 dB	30 dB	20 dB	10 dB
TV(λ)	0.01	0.1	1.0	10.0	10.0
Proposed(β)	0.001	0.001	1.0	10.0	100.0

tive comparison with existing methods. We used simulation data that were synthetically generated by the compartment model and object masks of brain phantom [13, 14]. The size of sinogram and PET image were $(182, 182)$ and $(128, 128)$, respectively ($N_\mathrm{s} = 182 * 182$, $N_\mathrm{i} = 128 * 128$). In the simulated images, there are three homogeneous regions with different kinetic parameters as shown in the left panel of Figure 6. The size of time domain was $T = 30$. We added Poisson noises to the sinograms such that the signal-to-noise ratio between $\boldsymbol{Y}^*$ and $\boldsymbol{Y}$ became almost {40 dB, 30dB, 20dB, or 10dB}. For the proposed method, we set $R = 3$, and the CNN architecture was the same as the U-Net with sigmoid output, and its details are explained in the

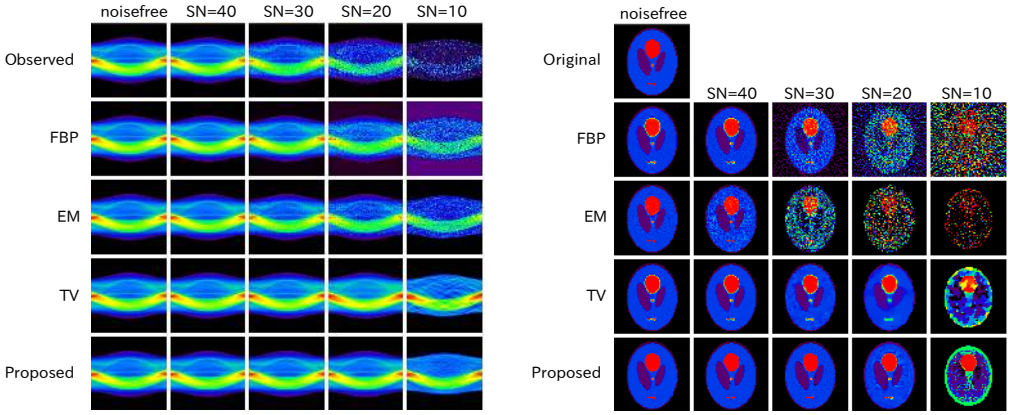


Figure 5. Comparison of reconstructed sinograms and PET images on brain phantom data with respect to SNR

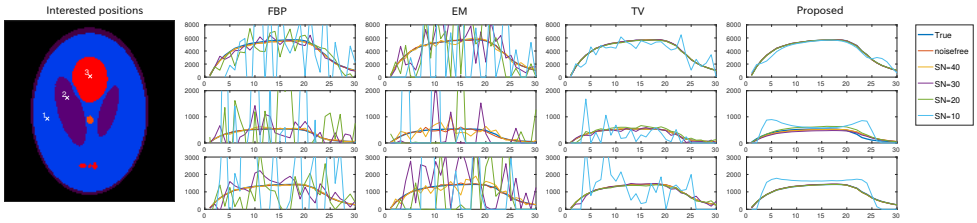


Figure 6. Comparison of reconstructed TACs. The TACs obtained at the locations 1, 2, and 3 indicated in the left panel are shown from top to bottom in the right panel. The horizontal axis of each graph of TACs shows the sampling index, $t = 1, 2, \ldots, T$.

supplemental materials.

4.1.1 Optimization behavior

We applied the proposed method to noisy sinograms. Figure 4 shows the decay of the cost function obtained when SNR=20dB, changes of spatial bases, the reconstructed sinogram, the reconstructed PET image, and its accuracy. The cost function continuously decreased with increasing iterations; and reconstruction accuracy of the sinogram and PET image showed peaks at 5,000 iterations. This is because that the spatial patterns overfitted the noisy sinogram.

4.1.2 Quantitative comparison with existing methods

PET image reconstruction from noisefree and noisy sinograms were conducted using the proposed method and three existing methods: FBP, EM, and TV regularization. As mentioned, both FBP and EM are the most widely employed methods for PET image reconstruction. We used a ramp filter for the FBP, which is the most standard. For EM, we just employed the results after optimization without filter. For TV regularization, we tried various values of $\lambda \in \{0.0001, 0.001, 0.01, 0.1, 1.0, 10.0, 100.0\}$ for all noisy sinograms and selected the best results. For the proposed method, we tried various values of $\beta \in \{0.0001, 0.001, 0.01, 0.1, 1.0, 10.0, 100.0\}$ for all noisy sinograms and selected the best results.

Figure 5 shows the visual comparison of all methods. FBP and EM were very sensitive to noise. The TV prior reconstructed smooth images but sometimes corrupted the detailed textures of the spatial domain. The proposed method reconstructed natural-looking images while keeping detailed textures. Figure 6 shows the TACs of three interested voxels. FBP and EM were largely vibrated because of noise. The TV prior for the spatial domain helps to smooth the time domain but did not smooth TACs well in the case of SNR=10dB. The proposed method reconstructed the smooth TACs.

Tables 1 and 2 show the values of SNR among the noisefree sinogram/PET image and the reconstructed sinograms/PET images. For this quantitative evaluation, EM results outperformed all other methods in the noise-free case; this is clearly true in theory. In noisy cases, the proposed method outperformed all other methods in PET image reconstruction. Table 3 shows the selected values of parameters for TV regularization and for the proposed method.

4.1.3 Blind decomposition under noisy measurement

Figure 7 shows the resultant spatio-temporal bases obtained from the phantom simulation data. The three homogeneous regions were successfully extracted from the noisy sinograms when the SNR was higher than 20dB. The ho-

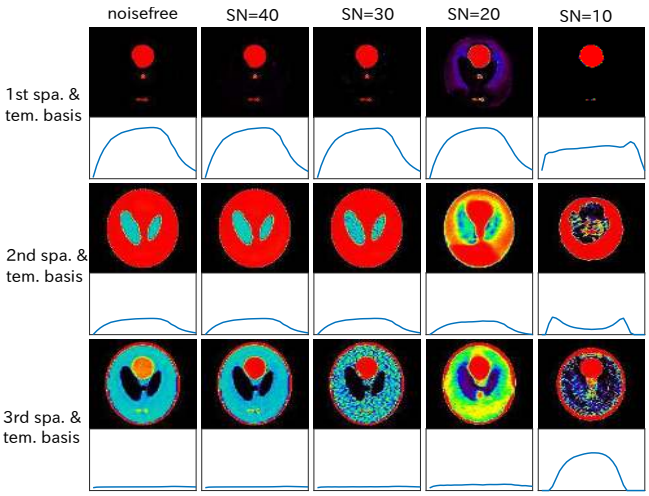


Figure 7. Spatio-temporal factors, $(\boldsymbol{a}_1, \boldsymbol{b}_1)$, $(\boldsymbol{a}_2, \boldsymbol{b}_2)$, and $(\boldsymbol{a}_3, \boldsymbol{b}_3)$, obtained by the proposed method. In each row, the spatial factor, $\boldsymbol{a}_r$, is shown at the top and the corresponding temporal factor, $\boldsymbol{b}_r$, is shown at the bottom.

mogeneous region extraction was difficult in the case of SNR=10dB.

5. Experiments using clinical data

In this section, we show the results of dynamic PET image reconstruction using clinical data. To evaluate the reconstruction performance with respect to different radioactive tracers, we used PET images captured with three different tracers, [^{11}C] carfentanil, [^{11}C] FLB 457, and [^{11}C] PiB. All these tracers indicate different biomolecular mechanisms. We obtained dynamic PET images of ten volunteers: four with PiB, three with carfentanil, and three with FLB 457. The sizes of the sinogram and PET image were $(128, 128)$ and $(128, 128)$, respectively. The size of the time domain was $T = 26$. We set $\lambda = 0.1$ for TV, and $\beta = 10.0$ for the proposed method.

Figure 8 shows the reconstructed sinograms and PET images by the proposed and three existing methods: the FBP, EM-based, and TV-regularization-based methods. FBP reconstructed the PET images, including some strong streak artifacts. EM reconstructed PET images that without streak artifacts but still with noise. TV reconstructed the smooth images; however, the contrast was decreased compared with the FBP and EM methods. The proposed method reconstructed high-contrast PET images with more piece-wise-smooth features and was the best according to a visual comparison.

Figure 9 shows examples of the spatio-temporal factors obtained from the dynamic PET images captured with a specific tracer, [^{11}C] FLB 457. It is known that [^{11}C] FLB 457 is uptaken strongly by the brain basal ganglia; two bright regions are located at the basal ganglia of the brain in each of the first spatial factors. It should be noted that these results were obtained without referring to any images from modalities that can capture anatomical shapes or structures such as the MRI [18], and that these regions, such as the brain basal ganglia, were detected by our method without using any knowledge of the brain's anatomy.

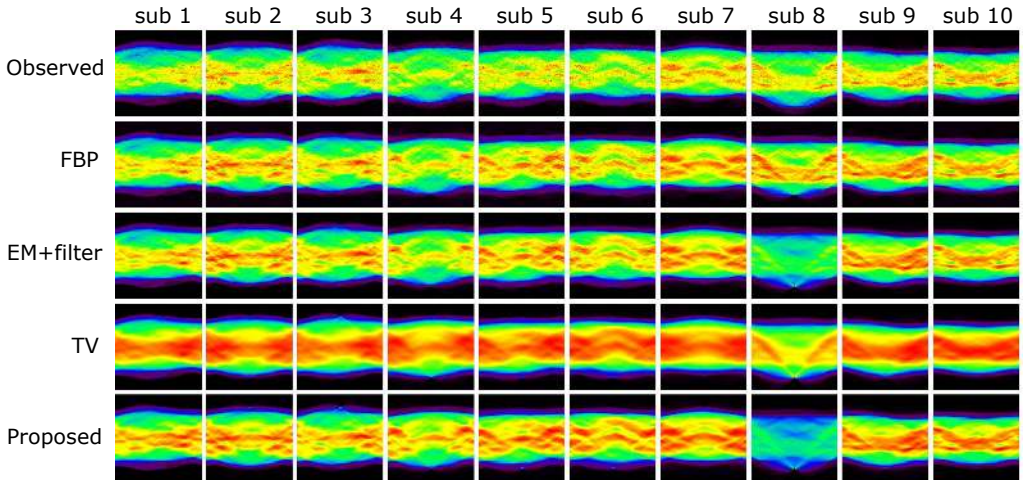
6. Discussions

As described in section 3.1, given a noise-free piece-wise constant PET image as a target, a randomly initialized single U-Net often outputs corresponding edge-enhanced images before the coefficient update process converges. Edge enhancement by a U-Net would be appealing for natural image retrieval but is not appropriate for medical image reconstruction from a measurement point of view. This is because edge enhancement degrades the SNR of images. Combining CNNs in parallel, we improved the output image quality: the step edges were well represented by simply adding all of the CNN outputs. The proposed method uses multiple CNNs, each of which generates each spatial factor, $\boldsymbol{a}_r$. Incorporating the CNNs into the NMF framework, the CNNs would output sparse spatial factors; this helps the separation of the homogeneous tissues with different kinetic parameters.

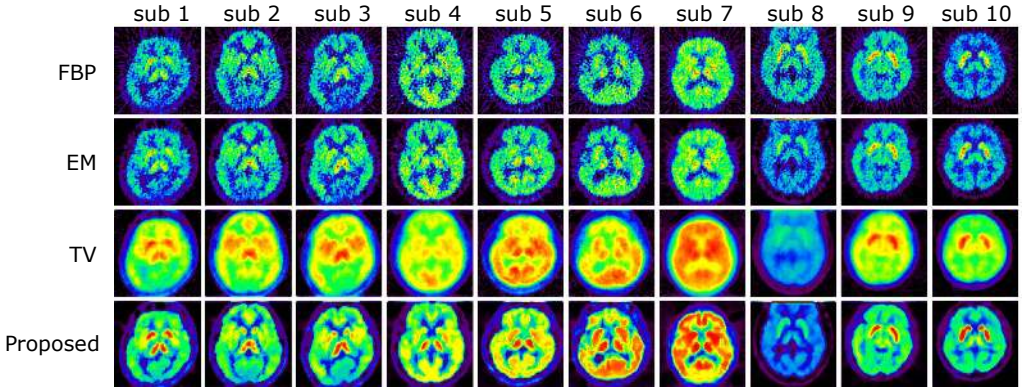
As shown in the experimental results, DIP is useful not only for generating natural pre-images [30] but also for reconstructing dynamic PET images from sinograms. In order to restrict the solution space of Problem (8), we must stop updating the coefficients, θ_r, of the CNNs before the convergence. Otherwise, the CNNs, $\phi(u|\theta_r)$, output images that are overfitted to the given patterns and that do not appropriately constrain the spatial factors. The maximum number of updates was determined empirically. It remains for a future work to develop a method of adaptively determining the number of parameter updates. Evaluating the change in image quality, such as the TV-norm, at each time the parameters are updated would be useful in determining the timing to stop the update of the parameters, θ_r. The proposed method optimizes not only $\boldsymbol{A}$ but also $\boldsymbol{Z}$. The stop time of the update should also be determined by changes in reconstructed dynamic images, $\boldsymbol{Z}$.

7. Conclusions

We proposed a method that reconstructs dynamic PET images from given sinograms by using NMF incorporated with DIP for appropriately constraining the spatial patterns of resultant image factors. The method reconstructs dynamic PET images with higher SNR and outputs pairs of spatial and temporal factors: the spatial factors represent homogeneous tissues with different kinetic parameters and the temporal factors represent the TACs that are observed in the corresponding homogeneous tissues. Experimen-



(a) Sinograms



(b) Reconstructed PET images

Figure 8. Comparison of reconstructed sinograms and PET images. Each column shows the results obtained from different volunteers: Three tracers, [^{11}C] carfentanil (Subjects 1-3), [^{11}C] PiB (Subjects 4-7) and [^{11}C] FLB 457 (Subjects 8-10), were used. From $T = 26$ frames, the sinogram and corresponding PET image with the highest average voxel value are shown.

tal results showed that the proposed method outperformed conventional methods and obtained the independent spatio-temporal factors, successfully. In clinical data, the reconstructed images are reasonable, and the region of basal ganglia was clearly extracted as spatial factor from the data of [^{11}C] FLB 457. Future works include evaluating in detail the images reconstructed by our method from a radiology point of view. This could be done to develop a method of determining the timing to stop updating the coefficients of the neural networks used for DIP.

Acknowledgement

This work was supported by JST ACT-I: Grant Number JPMJPR18UU, THE HORI SCIENCES AND ARTS FOUNDATION, and JSPS Grant-in-Aid for Scientific Research on Innovative Area (Multidisciplinary Computational Anatomy) JSPS KAKENHI: Grant Number 26108003. I'd like to thank Mr. Tomoshige Shimomura for his technical assistance.

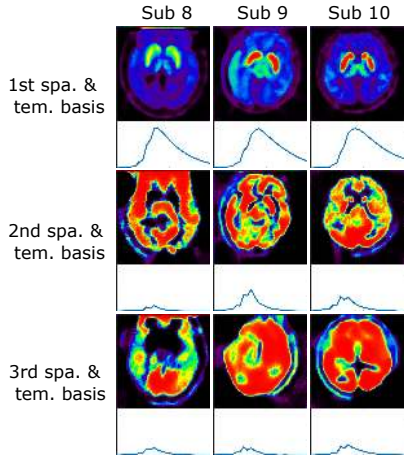


Figure 9. Spatio-temporal factors, $(\boldsymbol{a}_1, \boldsymbol{b}_1)$, $(\boldsymbol{a}_2, \boldsymbol{b}_2)$, and $(\boldsymbol{a}_3, \boldsymbol{b}_3)$, obtained from clinical data by the proposed method. Each column shows results obtained from different volunteer. In each row, the spatial factor, $\boldsymbol{a}_r$, is shown at the top and the corresponding temporal factor, $\boldsymbol{b}_r$, is shown at the bottom.

References

[1] Harrison H Barrett and William Swindell. *Radiological imaging: the theory of image formation, detection, and processing*, volume 2. Academic Press, 1996. 2

[2] Maitine Bergounioux, Evangelos Papoutsellis, Simon Stute, and Clovis Tauber. Infimal convolution spatiotemporal PET reconstruction using total variation based priors. preprint, 2018. 2

[3] Stephen Boyd, Neal Parikh, Eric Chu, Borja Peleato, and Jonathan Eckstein. Distributed optimization and statistical learning via the alternating direction method of multipliers. *Foundations and Trends in Machine Learning*, 3(1):1–122, 2011. 3

[4] Tony F Chan and Chiu-Kwong Wong. Total variation blind deconvolution. *IEEE Transactions on Image Processing*, 7(3):370–375, 1998. 2

[5] Laurent Condat. A primal–dual splitting method for convex optimization involving Lipschitzian, proximable and linear composite terms. *Journal of Optimization Theory and Applications*, 158(2):460–479, 2013. 3

[6] Sara Garbarino, Valentina Vivaldi, Fabrice Delbary, Giacomo Caviglia, Michele Piana, Cecilia Marini, Selene Capitanio, Iolanda Calamia, Ambra Buschiazzo, and Gianmario Sambuceti. A new compartmental method for the analysis of liver FDG kinetics in small animal models. *EJNMMI research*, 5(1):35, 2015. 2

[7] Kuang Gong, Ciprian Catana, Jinyi Qi, and Quanzheng Li. Pet image reconstruction using deep image prior. *IEEE Transactions on Medical Imaging*, 2018. 2

[8] Christine Guillemot and Olivier Le Meur. Image inpainting: Overview and recent advances. *IEEE Signal Processing Magazine*, 31(1):127–144, 2014. 2

[9] Jieqing Jiao, Alexandre Bousse, Kris Thielemans, Ninon Burgos, Philip SJ Weston, Jonathan M Schott, David Atkinson, Simon R Arridge, Brian F Hutton, Pawel Markiewicz, et al. Direct parametric reconstruction with joint motion estimation/correction for dynamic brain PET data. *IEEE Transactions on Medical Imaging*, 36(1):203–213, 2017. 1

[10] Mustafa E Kamasak, Charles A Bouman, Evan D Morris, and Ken Sauer. Direct reconstruction of kinetic parameter images from dynamic PET data. *IEEE Transactions on Medical Imaging*, 24(5):636–650, 2005. 1

[11] Kazuya Kawai, Hidekata Hontani, Tatsuya Yokota, Muneyuki Sakata, and Yuichi Kimura. Simultaneous PET image reconstruction and feature extraction method using non-negative, smooth, and sparse matrix factorization. In *Asia-Pacific Signal and Information Processing Association Annual Summit and Conference (APSIPA ASC), 2018*, pages 1334–1337. IEEE, 2018. 2, 3

[12] Kazuya Kawai, Junya Yamada, Hidekata Hontani, Tatsuya Yokota, Muneyuki Sakata, and Yuichi Kimura. A robust PET image reconstruction using constrained non-negative matrix factorization. In *Asia-Pacific Signal and Information Processing Association Annual Summit and Conference (APSIPA ASC), 2017*, pages 1815–1818. IEEE, 2017. 2, 3

[13] AA Lammertsma, CJ Bench, SP Hume, S Osman, K Gunn, DJ Brooks, and RSJ Frackowiak. Comparison of methods for analysis of clinical [11C] raclopride studies. *Journal of Cerebral Blood Flow & Metabolism*, 16(1):42–52, 1996. 5

[14] Adriaan A Lammertsma and Susan P Hume. Simplified reference tissue model for PET receptor studies. *Neuroimage*, 4(3):153–158, 1996. 5

[15] Kenneth Lange, Mark Bahn, and Roderick Little. A theoretical study of some maximum likelihood algorithms for emission and transmission tomography. *IEEE Transactions on Medical Imaging*, 6(2):106–114, 1987. 3

[16] Daniel D Lee and H Sebastian Seung. Learning the parts of objects by non-negative matrix factorization. *Nature*, 401(6755):788, 1999. 2, 3

[17] Daniel D Lee and H Sebastian Seung. Algorithms for non-negative matrix factorization. In *Advances in Neural Information Processing Systems*, pages 556–562, 2001. 2, 3

[18] Rebekka Loeb, Nassir Navab, and Sibylle I Ziegler. Direct parametric reconstruction using anatomical regularization for simultaneous PET/MRI data. *IEEE Transactions on Medical Imaging*, 34(11):2233–2247, 2015. 7

[19] Jean Logan. Graphical analysis of PET data applied to reversible and irreversible tracers. *Nuclear Medicine and Biology*, 27(7):661–670, 2000. 1

[20] Jean Logan, Joanna S Fowler, Nora D Volkow, Alfred P Wolf, Stephen L Dewey, David J Schlyer, Robert R MacGregor, Robert Hitzemann, Bernard Bendriem, S John Gatley, et al. Graphical analysis of reversible radioligand binding from timeactivity measurements applied to [N-11C-methyl]-(-)-cocaine PET studies in human subjects. *Journal of Cerebral Blood Flow & Metabolism*, 10(5):740–747, 1990. 1

[21] Steven R Meikle, Julian C Matthews, Vincent J Cunningham, Dale L Bailey, Lefteris Livieratos, Terry Jones, and Pat Price. Parametric image reconstruction using spectral analysis of PET projection data. *Physics in Medicine & Biology*, 43(3):651, 1998. 2

[22] Thibaut Merlin, Dimitris Visvikis, Philippe Fernandez, and Frédéric Lamare. Dynamic PET image reconstruction integrating temporal regularization associated with respiratory motion correction for applications in oncology. *Physics in Medicine & Biology*, 63(4):045012, 2018. 2

[23] Frank Natterer. *The Mathematics of Computerized Tomography*, volume 32. Siam, 1986. 2

[24] Arman Rahmim, Jing Tang, and Habib Zaidi. Four-dimensional (4D) image reconstruction strategies in dynamic PET: Beyond conventional independent frame reconstruction. *Medical Physics*, 36(8):3654–3670, 2009. 2

[25] Andrew J Reader, Julian C Matthews, Florent C Sureau, Claude Comtat, Régine Trebossen, and Irène Buvat. Iterative kinetic parameter estimation within fully 4D PET image reconstruction. In *Nuclear Science Symposium Conference Record, 2006. IEEE*, volume 3, pages 1752–1756. IEEE, 2006. 1

[26] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. U-Net: Convolutional networks for biomedical image segmentation. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 234–241. Springer, 2015. 4

[27] Mara Scussolini, Sara Garbarino, Gianmario Sambuceti, Giacomo Caviglia, and Michele Piana. A physiology-based

parametric imaging method for FDG–PET data. *Inverse Problems*, 33(12):125010, 2017. 1, 2

[28] Lawrence A Shepp and Yehuda Vardi. Maximum likelihood reconstruction for emission tomography. *IEEE Transactions on Medical Imaging*, 1(2):113–122, 1982. 1, 3

[29] Feng Shi, Jian Cheng, Li Wang, Pew-Thian Yap, and Dinggang Shen. LRTV: MR image super-resolution with low-rank and total variation regularizations. *IEEE Transactions on Medical Imaging*, 34(12):2459–2466, 2015. 2

[30] Dmitry Ulyanov, Andrea Vedaldi, and Victor Lempitsky. Deep image prior. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2018. 2, 3, 7

[31] Richard J Walledge, Roido Manavaki, Michael Honer, and Andrew J Reader. Inter-frame filtering for list-mode em reconstruction in high-resolution 4-d PET. *IEEE Transactions on Nuclear Science*, 51(3):705–711, 2004. 2

[32] Hiroshi Watabe, Yoko Ikoma, Yuichi Kimura, Mika Naganawa, and Miho Shidahara. PET kinetic analysiscompartmental model. *Annals of Nuclear Medicine*, 20(9):583, 2006. 1, 2

[33] Xingjian Yu, Chenye Wang, Hongjie Hu, and Huafeng Liu. Low dose PET image reconstruction with total variation using alternating direction method. *PloS one*, 11(12):e0166871, 2016. 2