Semi-Supervised Tripled Dictionary Learning for Standard-dose PET Image Prediction using Low-dose PET and Multimodal MRI

Yan Wang,

College of Computer Science, Sichuan University, Chengdu, China

Guangkai Ma,

Department of Radiology and BRIC, University of North Carolina at Chapel Hill, NC, USA

Le An,

Department of Radiology and BRIC, University of North Carolina at Chapel Hill, NC, USA

Feng Shi,

Department of Radiology and BRIC, University of North Carolina at Chapel Hill, NC, USA

Pei Zhang,

Department of Radiology and BRIC, University of North Carolina at Chapel Hill, NC, USA

David S. Lalush,

Joint Department of Biomedical Engineering, University of North Carolina at Chapel Hill and North Carolina State University, NC, USA

Xi Wu,

Department of Computer Science, Chengdu University of Information Technology, Chengdu, China

Yifei Pu,

College of Computer Science, Sichuan University, Chengdu, China

Jiliu Zhou, and

College of Computer Science, Sichuan University, Chengdu, ChinaDepartment of Computer Science, Chengdu University of Information Technology, Chengdu, China

Dinggang Shen[*]

Department of Radiology and BRIC, University of North Carolina at Chapel Hill, NC, USA

Abstract

Objective—To obtain high-quality positron emission tomography (PET) image with low-dose tracer injection, this study attempts to predict the standard-dose PET (S-PET) image from both its low-dose PET (L-PET) counterpart and corresponding magnetic resonance imaging (MRI).

Methods—It was achieved by patch-based sparse representation (SR), using the training samples with a complete set of MRI, L-PET and S-PET modalities for dictionary construction. However,

[*]Correspondence: dgshen@med.unc.edu.

the number of training samples with complete modalities is often limited. In practice, many samples generally have incomplete modalities (i.e., with one or two missing modalities) that thus cannot be used in the prediction process. In light of this, we develop a semi-supervised tripled dictionary learning (SSTDL) method for S-PET image prediction, which can utilize not only the samples with complete modalities (called complete samples) but also the samples with incomplete modalities (called incomplete samples), to take advantage of the large number of available training samples and thus further improve the prediction performance.

Results—Validation was done on a real human brain dataset consisting of 18 subjects, and the results show that our method is superior to the SR and other baseline methods.

Conclusion—This work proposed a new S-PET prediction method, which can significantly improve the PET image quality with low-dose injection.

Significance—The proposed method is favorable in clinical application since it can decrease the potential radiation risk for patients.

Index Terms

Positron emission tomography (PET); Sparse representation (SR); Local coordinate coding (LCC); Semi-supervised tripled dictionary learning (SSTDL)

I. Introduction

Nowadays, positron emission tomography (PET) is increasingly and widely used in hospitals and clinics for disease diagnosis and intervention. Different from computed tomography (CT) and magnetic resonance imaging (MRI), PET provides insight into the biochemical and physiological processes of the human body [1]. Due to its unique advantages, PET has been widely used in many medical imaging applications, such as clinical oncology [2], cardiac usages [3], and certain brain diseases [4–6]. The PET scanning is non-invasive, however, the radiotracer used for PET imaging (e.g., 18F-FDG) involves ionizing radiation. The standard-dose scans in our cohort averaged 203 MBq of 18F-FDG, which corresponds to an effective dose of 3.86 mSv. Based on the report "Biological Effects of Ionizing Radiation (BEIR VII)", the increased risk of incidence of cancer is 10.8% per Sv, so one brain PET scan increases lifetime cancer risk by about 0.04%. The International Commission on Radiation Protection considers the increased risk of death by cancer to be about 4% per Sv, so one brain PET scan increases the risk of death from cancer by 0.015%. These numbers are small, but the risks are multiplied for patients who undergo multiple PET scans as part of their treatment regimen. In addition, pediatric patients have increased risks. Therefore, the long-term focus of this work is to reduce the total dose for these non-standard populations at increased risk from PET radiation dose. However, reducing the tracer dose will degrade the PET image quality, since PET imaging is a quantum accumulation process. Hence, to estimate the high-quality standard-dose PET (S-PET) image from a low-dose PET (L-PET) image is a promising alternative approach and is of great research interest. Moreover, it has been shown that MR images can be used to improve the PET reconstruction image. Especially in PET brain imaging, the added value of using MR images during PET reconstruction (i.e., as anatomical prior) or after PET reconstruction (i.e., for partial volume correction) has been studied in great detail already over the past decades. A recent review

discussing these methods can be found in [7]. In this paper, we explore the use of both L-PET image and multimodal MRI (T1-weighted and diffusion tensor imaging (DTI)) to estimate the S-PET image. Note that the common DTI measures include fractional anisotropy (FA), mean diffusivity (MD), radial diffusivity (RD), etc. Here, we compute FA and MD images from diffusion images for S-PET image prediction.

Sparse representation (SR) is an approach to represent a given signal as a linear combination of a small number of elements in a dictionary, and it has been widely used in different areas of image processing, such as image resolution enhancement [8], and image denoising [9]. Although the real-world PET images are often not sparse, they could be approximated by their sparse components, leading to the application of SR technique on PET images. Many studies have shown that the performance of SR depends on not only the signal itself but also the completeness of the dictionary [10]. Various approaches have been proposed to build over-complete dictionaries. For SR application in resolution enhancement, the dictionaries are built with all training samples [11]. Specifically, the sparse coefficients estimated from the low-resolution dictionary (constructed by all the low-resolution training image patches) are applied to the corresponding high-resolution dictionary (constructed by all the corresponding high-resolution training image patches). In this way, the high-resolution image patch is reconstructed from a given low-resolution image patch. This kind of dictionary construction approach typically requires coupled samples, i.e., each sample in the training set should have both low-resolution and high-resolution images. Similarly, following the same idea as in image super-resolution, PET prediction requires each sample in the training set to have the complete modalities, including multimodal MR (T1, FA and MD), L-PET and S-PET images. Nevertheless, this requirement is hardly satisfied in practice. To address this issue, we propose to effectively utilize all the available training samples (referred to as incomplete) to predict the S-PET image.

Instead of using predefined dictionaries, dictionary learning (DL) has been extensively studied to learn the dictionary atoms that are adapted to the data with a specific distribution. The DL technique has been widely applied in PET image analysis. Specifically, authors in [12] proposed an adaptive dictionary learning approach for PET image deblurring while suppressing Poisson noise effects. In [13], a reconstruction framework integrating SR and DL into a maximum likelihood estimator has been proposed for accurate and robust PET image reconstruction. Besides, dual dictionary learning has been successfully applied to biomedical imaging as well [14–16]. Recently, local coordinate coding (LCC) based dictionary learning, which can find non-zero coefficients for dictionary atoms that are neighbors of the target sample for SR, has shown promising results in many applications [17–21]. Inspired by LCC, we propose an efficient semi-supervised tripled dictionary learning method to predict the S-PET image from L-PET and multimodal MRI. Also, before dictionary learning, a graph-based mapping strategy is developed to make the embedded geometric relationship of the patches in the multimodal MRI/L-PET similar to the relationship of those in the S-PET. In addition, a fusion strategy for multimodal MRI is further developed to avoid any prediction bias due to the possible domination of multimodal MRI over the L-PET, and also to balance the contribution of each channel of multimodal MRI in SR. After mapping and multimodal MRI fusion, the fused multimodal MRI, L-PET and S-PET will have similar data distributions (i.e., similar embedded geometric relationship

of the patches). Then, in the training stage, we use the complete samples to explore the relationship between MRI/L-PET and S-PET patches. This is to ensure that the SR of MRI and L-PET patches using the MRI and L-PET dictionaries can be directly used to reconstruct the corresponding S-PET patch using the S-PET dictionary. In addition, the incomplete samples are introduced to ensure that each learned dictionary can also well reconstruct patches from each modality. In the testing stage, given a sample with multimodal MRI and L-PET images, its coupled MRI and L-PET patches can be jointly encoded by the learned MRI and L-PET dictionaries via LCC, to obtain the sparse coefficients. Finally, the prediction can be performed by applying the obtained sparse coefficients to the learned S-PET dictionary.

Below, we first introduce the preprocessing procedure, including both the graph-based mapping strategy and the multimodal MRI fusion strategy, in Section 2. Then, Section 3 presents the details of semi-supervised tripled dictionary learning based on LCC with an instantiation of our problem. Finally, the experimental results are shown in Section 4, and conclusions are drawn in Section 5.

II. Data Preprocessing: Mapping and Fusion

Suppose that the training set consists of the complete samples with multimodal MRI (T1, FA and MD), L-PET and S-PET images, as well as the incomplete samples missing images from one or more modalities. Before dictionary learning, two procedures will be performed, including the graph-based data mapping and the multimodal MRI fusion. The motivation and technical details are discussed in the following subsections.

A. Graph-based Data Mapping

The quality of the predicted S-PET image largely depends on whether the learned sparse representation can well reconstruct the testing image. By applying the traditional SR, the underlying assumption is that the embedded geometric relationship of patches in the S-PET image space (i.e., represented by graph g^S) is very similar to that of patches in the MRI/L-PET image space (i.e., represented by graphs g^{MR}/g^L). Note that a graph here represents the feature distribution of image patches. Specifically, each node in the graph represents a patch, and each edge describes the geometric relationship between a pair of patches. Based on the above assumption, the sparse coefficients estimated from the MRI/L-PET patch space can be directly used to predict the S-PET image patches in the S-PET patch space. However, the above assumption may not be satisfied due to the huge difference in imaging mechanisms between PET and MR images, which is further enlarged by different imaging noise introduced during acquisition. Therefore, to make the geometric relationship of patches between MRI/L-PET and S-PET more similar, we should find a way to transform the graphs g^{MR}/g^L MRI/L-PET images to best match the graph g^S of the respective S-PET image. This can be done by enforcing node-to-node matching, edge-to-edge matching, or even high-order matching between the g^{MR}/g^L and g^S In this study, we consider only the node-to-node and edge-to-edge matching, although the plane-to-plane matching or higher-order matching can be easily realized.

Taking the mapping between L-PET and S-PET images for example, we first find the samples with both L-PET and S-PET images from the training set, and then compute their mapping matrix M^L as follows:

$$\underset{M^L}{\arg\min} \sum_i \left\| \boldsymbol{x}_i^S - M^L * \boldsymbol{x}_i^L \right\|_2^2$$
$$+ \gamma \sum_i \sum_j \left\| (\boldsymbol{x}_i^S - \boldsymbol{x}_j^S) - M^L * (\boldsymbol{x}_i^L - \boldsymbol{x}_j^L) \right\|_2^2 \quad (1)$$

where $\boldsymbol{x}_i^S$ is an S-PET image patch, $\boldsymbol{x}_i^L$ is its corresponding L-PET image patch, and M^L is a mapping matrix to transform $\boldsymbol{x}_i^L$ Eq. (1) is a least-square problem and it has an analytical solution. The patch size in this paper is set to be 5×5×5. In our case, we reshape the patch into a vector with 125 elements, and thus the dimension of the mapping matrix M^L is 125×125. The first term represents node-to-node matching between graphs g^S and g^L; that is, the mapped L-PET image patch $M^L * \boldsymbol{x}_i^L$ should be similar to $\boldsymbol{x}_i^S$. The second term enforces that the relationship between the mapped image patches $M^L \times \boldsymbol{x}_i^L$ and $M^L \times \boldsymbol{x}_j^L$ should be very similar to that between $\boldsymbol{x}_i^S$ and $\boldsymbol{x}_j^S$, which can be regarded as edge-to-edge matching. Examples of node-to-node matching, edge-to-edge matching, and high-order matching are given in Fig. 1. In this paper, index i of Eq. (1) is running over all patches of the training images to ensure the node-to-node matching. For reducing the computational complexity, index j is just running over the patches within a neighborhood (with a size of 15×15×15) centered at the voxel assigned with index i across all training images, to enforce the edge-to-edge matching. After solving the above equation and obtaining the mapping matrix M^L, the mapped L-PET image patch $M^L * \boldsymbol{x}_j^L$ can be easily calculated. The mapping matrices M^{T1}, M^{FA}, M^{MD} between MRI (T1, FA and MD) and S-PET images can be estimated in a similar way, and then the mapped T1, FA and MD image patches can be obtained accordingly (see Fig. 2).

B. Multimodal MRI Fusion

Since we employ three MRI modalities (T1, FA and MD, with the latter two computed from the same DTI) for helping S-PET image prediction, it is important to combine the useful information from different MRI modalities into a single MRI representation. To this end, we unify three different MRI patches by weighted averaging. To determine the weight for each modality, the following optimization problem is employed:

$$\underset{\mathbf{w}}{\arg\min} \sum_i \left\| \boldsymbol{x}_i^S - (w_{T_1} M^{T1} \boldsymbol{x}_i^{T1} + w_{FA} M^{FA} \boldsymbol{x}_i^{FA} \right.$$
$$\left. + w_{MD} M^{MD} \boldsymbol{x}_i^{MD}) \right\|$$
$$\text{s. t. } w_{T_1} + w_{FA} + w_{MD} = 1 \quad (2)$$

where $\boldsymbol{x}_i^{T1}$, $\boldsymbol{x}_i^{FA}$, and $\boldsymbol{x}_i^{MD}$ represent the T1, FA and MD image patches, respectively, $\mathbf{w} = w_{T1}, w_{FA}, w_{MD}]$ is a vector of three weights associated with $\boldsymbol{x}_i^{T1}$, $\boldsymbol{x}_i^{FA}$, and $\boldsymbol{x}_i^{MD}$,

respectively. Eq. (2) is a constrained least-square problem and has an analytical solution. Again, the index i is running over all patches of the training images, which has the same meaning as in Eq. (1). Then, with these estimated weights, the fused image patch $\boldsymbol{x}_i^{MR}$ from multimodal MRI patches can be obtained as follows:

$$\boldsymbol{x}_i^{MR} = w_{T1} M^{T1} \boldsymbol{x}_i^{T1} + w_{FA} M^{FA} \boldsymbol{x}_i^{FA} + w_{MD} M^{MD} \quad (3)$$

The mapping and fusion procedures of multimodal MR images are illustrated in Fig. 2.

All mapped L-PET patches as well as fused MR patches will be used for subsequent dictionary learning. For simplicity, in the following, we will use the L-PET image to refer to the mapped L-PET image, and the MR image to refer to the fused image from the mapped T1, FA and MD images.

III. Semi-supervised Tripled Dictionary Learning

In this section, we first briefly introduce the LCC method that serves as the basis of the dictionary learning in our method, and then the details of the proposed semi-supervised tripled dictionary learning method will be discussed.

A. LCC based Dictionary Learning

LCC based dictionary learning aims to learn a dictionary that can best reconstruct a (training) patch $\boldsymbol{x}_i$ locality during coding [22]. Here, the locality means that the patches are represented by their neighboring atoms in the dictionary. Specifically, given N H-dimensional training patches $[\boldsymbol{x}_1, \ldots, \boldsymbol{x}_i, \ldots, \boldsymbol{x}_N] \in \mathbb{R}^{H \times N}$ (where H indicates the patch size, and N is the number of patches in each image times the number of training images), the dictionary $D = [\boldsymbol{d}_1, \ldots, \boldsymbol{d}_q, \ldots, \boldsymbol{d}_P] \in \mathbb{R}^{H \times P}$ (where P indicates the number of the atoms in the dictionary) can be learned by minimizing the following objective function:

$$\underset{D, \{\boldsymbol{\alpha}_i\}}{\arg\ \min} LCC(D, \{\boldsymbol{\alpha}_i\}, \{\boldsymbol{x}_i\}) =$$
$$\underset{D, \{\boldsymbol{\alpha}_i\}}{\arg\ \min} \sum_i \left(\frac{1}{2} \|\boldsymbol{x}_i - D\boldsymbol{\alpha}_i\|^2 + \mu \sum_q |\alpha_i^q| \|\boldsymbol{d}_q - \boldsymbol{x}_i\|^2 \right) \quad (4)$$

where $\boldsymbol{x}_i$ is the i-th patch to be represented, $\boldsymbol{a}_i = [\alpha_i^1, \ldots, \alpha_i^q, \ldots, \alpha_i^p]^T$ is the corresponding sparse coefficient of $\boldsymbol{x}_i$, α_i^q is the q-th component of $\boldsymbol{a}_i$, and $D = [\boldsymbol{d}_1, \ldots, \boldsymbol{d}_q, \ldots, \boldsymbol{d}_p]$ is the dictionary to be learned with $\boldsymbol{d}_q$ as the q-th column of the dictionary D. The first term measures the reconstruction error, while the second term preserves the locality of the coding by strongly penalizing dictionary atoms that are far from $\boldsymbol{x}_i$. The parameter μ is used to balance the construction error and locality penalty.

Although the above objective function is not jointly convex over D and $\{\boldsymbol{a}_i\}$ it is convex with respect to each unknown variable when the other one is fixed. As a result, Eq. 4 can be alternatively optimized by updating one variable at a time. Using this strategy, the

optimization problem of the above objective function can be divided into two sub-problems as detailed below: 1) updating sparse coefficients {$\boldsymbol{a}_i$} while fixing D, and 2) updating dictionary D while fixing {$\boldsymbol{a}_i$}.

Updating Sparse Coefficients—We first fix the dictionary D to update the sparse coefficient vector {$\boldsymbol{a}_i$}, which can be converted to a sparse coding problem. To initialize the dictionary, each atom is generated by performing random linear combination of all N training patches, with the sum of combination weights being equal to 1. Then, let $\boldsymbol{\beta}_i = \Lambda_i \boldsymbol{a}_i$ where $\Lambda_i = diag(\lambda_i^1, \ldots, \lambda_i^q, \ldots, \lambda_i^P)$ is a diagonal matrix whose diagonal elements are $\lambda_i^q = \left\| \boldsymbol{d}_q - \boldsymbol{x}_i \right\|^2$ Using the above transformation, the objective function of LCC based dictionary learning can be expressed as a LASSO problem [23]:

$$
\underset{\{\boldsymbol{\alpha}_i\}}{\arg\min} LCC(D, \{\boldsymbol{\alpha}_i\}, \{\boldsymbol{x}_i\}) = \\
\underset{\boldsymbol{\beta}_i}{\arg\min} \sum_i \left(\frac{1}{2} \|\boldsymbol{x}_i - D\Lambda_i^{-1}\boldsymbol{\beta}_i\|^2 + \mu \|\boldsymbol{\beta}_i\|_1 \right) \quad (5)
$$

where $\left\| \boldsymbol{\beta}_i \right\|_1 = \sum_q |\beta_i^q|$ denotes the l_1-norm. After $\boldsymbol{\beta}_i$ is solved, $\boldsymbol{a}_i$ is obtained by $\boldsymbol{\alpha}_i = \Lambda_i^{-1}\boldsymbol{\beta}_i$.

Updating Dictionary—After solving {$\boldsymbol{a}_i$} for all patches, optimizing D becomes a constrained quadratic programming problem. Specifically, by expanding the squares of the objective function and omitting the terms without D, we can obtain:

$$
\underset{D}{\arg\min} LCC(D, \{\boldsymbol{\alpha}_i\}, \{\boldsymbol{x}_i\}) = \\
\underset{D}{\arg\min} \frac{1}{2} tr[D^T D(\sum_i \boldsymbol{\alpha}_i \boldsymbol{\alpha}_i^T + 2\mu \Psi_i)] \\
- tr[D^T(\sum_i \boldsymbol{x}_i \boldsymbol{\alpha}_i^T + 2\mu \boldsymbol{x}_i \overline{\boldsymbol{\alpha}}_i^T)] \quad (6)
$$

where $\overline{\boldsymbol{\alpha}}_i$ is component-wise absolute value of $\boldsymbol{a}_i$, i.e., $\overline{\alpha}_i^q = |\alpha_i^q|$, and Ψ_i is a diagonal matrix with its diagonal elements constructed from $\overline{\boldsymbol{\alpha}}_i$ i.e., $\Psi_i = diag(\overline{\boldsymbol{\alpha}}_i)$ and $tr[\cdot]$ represents the trace of the square matrix, which computes the sum of the elements on the diagonal of the square matrix. Let $E = \sum_i \boldsymbol{\alpha}_i \boldsymbol{\alpha}_i^T + 2\mu \Psi_i$, $F = \sum_i \boldsymbol{x}_i \boldsymbol{\alpha}_i^T + 2\mu \boldsymbol{x}_i \overline{\boldsymbol{\alpha}}_i^T$ and $\boldsymbol{e}_q$ and $\boldsymbol{f}_q$ represent the q-th columns of matrices E and F. Then, the optimization of D can be solved by using the gradient descent method. Specifically, in the k-th iteration, the q-th column of D, $\boldsymbol{d}_q^{(k)}$, can be updated as follows:

$$
\boldsymbol{d}_q^{(k+1)} = \boldsymbol{d}_q^{(k)} - \frac{1}{\Delta} \left(D^{(k)}\boldsymbol{e}_q - \boldsymbol{f}_q \right), \quad (7)
$$

where $D^{(k)}$ is the dictionary at the k-th iteration, and Δ is a scalar controlling the step size. Let N_A denote the number of. iterations for updating the atoms of dictionary.

The steps of 1) updating sparse coefficients and 2) updating dictionary iterate until convergence, in order to obtain the optimized dictionary. As the iteration increases, the objective function will gradually approach its optimum. The iteration will end when the difference between two iterations is lower than the predefined threshold or the iteration reaches the maximum number. In this paper, we set the number of iterations for the alternating optimization in dictionary learning, denoted as N_D, to be 20. This number was chosen, because it results in a very small change in objective function value between subsequent iterations, i.e., less than 0.01%, and can therefore be considered sufficient to lead to convergence.

B. Semi-Supervised Tripled Dictionary Learning for S-PET Image Prediction

In a supervised dictionary learning scheme, the complete samples with all MRI, L-PET and S-PET images are required for prediction of S-PET image. However, in practice, it is hard to simultaneously obtain all the modality images for a large number of subjects, while many subjects have only one or two modality image(s) (i.e., MRI, L-PET, or S-PET). In order to also utilize those incomplete samples for improving the dictionary learning, we propose a semi-supervised tripled dictionary learning (SSTDL) method that can utilize all the complete and incomplete data.

In the training stage, both complete and incomplete samples are used in a semi-supervised manner to learn three dictionaries, D_{MR}, D_L and D_S for MRI, L-PET and S-PET, respectively. In the testing stage, given an MR image and the associated L-PET image, the goal is to predict the target S-PET image by first jointly encoding the MR and L-PET images using the dictionaries D_{MR} and D_L and then applying the obtained sparse coefficients to the dictionary D_S The flowchart of the proposed semi-supervised tripled dictionary learning for prediction of S-PET image is shown in Fig. 3. In our method, we assume that, if atoms in the dictionary are similar to the patch of the testing voxel (under prediction), the related patches should also have similar S-PET values. Considering the smoothness of anatomical structures of human brain, for each testing voxel (under prediction), the patches near the testing voxel should be also similar. Therefore, to ensure the collection of enough atoms in the dictionary that are relevant to the testing voxel and also able to reconstruct this testing sample, we define a neighborhood centered at the testing voxel to extract image patches for learning the dictionaries. In the following subsection, the training stage of the SSTDL method is presented in detail.

1) Semi-Supervised Tripled Dictionary Learning

In our method, each voxel of a target S-PET image is predicted independently. Specifically, to predict the value of a voxel z in the unknown S-PET image, we first define a neighborhood centered at voxel z with the size $\omega \times \omega \times \omega$ in the training images. By grouping all the patches with the same size of $\rho \times \rho \times \rho$ centered at voxels in the same neighborhood across all MRI, L-PET and S-PET of training subjects, three patch sets are generated: $\{x_1^{MR}, \ldots, x_i^{MR}, \ldots, x_{N_{MR}}^{MR}\} \in \mathbb{R}^{H \times N_{MR}}$, $\{x_1^L, \ldots, x_i^L, \ldots, x_{N_L}^L\} \in \mathbb{R}^{H \times N_L}$, and $\{x_1^S, \ldots, x_i^S, \ldots, x_{N_S}^S\} \in \mathbb{R}^{H \times N_S}$, where $H = \rho \times \rho \times \rho$ is the dimensionality of the patch, N_{MR}, N_L and N_S are the numbers of MRI, L-PET and S-PET patches, respectively. Assuming that the top t tripled patches

$\{(\boldsymbol{x}_i^{MR}, \boldsymbol{x}_i^L, \boldsymbol{x}_i^S)\}_{i=1}^t (t \leq N_{MR}, t \leq N_L, \text{and } t \leq N_s)$ are acquired from the complete samples, and the other patches $\{\boldsymbol{x}_i^{MR}\}_{i=t+1}^{N_{MR}}, \{\boldsymbol{x}_i^L\}_{i=t+1}^{N_L}, \{\boldsymbol{x}_i^S\}_{i=t+1}^{N_S}$ are acquired from the incomplete samples, the goal is to learn three dictionaries D_{MR}, D_L and D_S, related to the MRI, L-PET and S-PET patch spaces, respectively. In the dictionary learning, each tripled patch $(\boldsymbol{x}_i^{MR}, \boldsymbol{x}_i^L, \boldsymbol{x}_i^S)$, $i = 1, \ldots, t$ of three modalities are represented by the common coefficient vector $\boldsymbol{\alpha}_i^{(C)}$ using the dictionaries D_{MR}, D_L and D_S. For incomplete data, $\{\boldsymbol{x}_i^{MR}\}_{i=t+1}^{N_{MR}}, \{\boldsymbol{x}_i^L\}_{i=t+1}^{N_L}, \{\boldsymbol{x}_i^S\}_{i=t+1}^{N_S}$ are respectively represented by the coefficient vectors $\{\boldsymbol{\alpha}_i^{(MR)}\}_{i=t+1}^{N_{MR}}, \{\boldsymbol{\alpha}_i^{(L)}\}_{i=t+1}^{N_L}$ and $\{\boldsymbol{\alpha}_i^{(S)}\}_{i=t+1}^{N_S}$ using D_{MR}, D_L and D_S, Let's set $A^{(C)} = \{\boldsymbol{\alpha}_i^{(C)}\}_{i=1}^t$, $A^{(MR)} = \{\boldsymbol{\alpha}_i^{(MR)}\}_{i=t+1}^{N_{MR}}$, $A^{(L)} = \{\boldsymbol{\alpha}_i^{(L)}\}_{i=t+1}^{N_L}$ and $A^{(S)} = \{\boldsymbol{\alpha}_i^{(S)}\}_{i=t+1}^{N_S}$.

In the proposed semi-supervised tripled dictionary learning (SSTDL) method, the tripled patches can be used to enforce that the common sparse representation $A^{(C)}$ of an MRI patch and its associated L-PET patch, using D_{MR} and D_L can well reconstruct their respective S-PET patch, using D_S. Patches representation ability of the dictionary, i.e., the learned dictionaries can also well construct the corresponding incomplete samples. By using both the complete and the incomplete samples, the objective function of SSTDL is given by:

$$\begin{aligned} &\Phi(D_{MR}, D_L, D_S, A^{(C)}, A^{(MR)}, A^{(L)}, A^{(S)}) = \\ &E_{complete}(D_{MR}, D_L, D_S, A^{(C)}) + E_{MR}(D_{MR}, A^{(MR)}) \\ &\qquad\qquad + E_L(D_L, A^{(L)}) + E_S(D_S, A^{(S)}) \end{aligned} \qquad (8)$$

The above objective function includes four terms: one complete term $E_{complete}$ and three incomplete terms E_{MR}, E_L and E_S. The complete term uses the tripled patches $\{(\boldsymbol{x}_i^{MR}, \boldsymbol{x}_i^L, \boldsymbol{x}_i^S)\}_{i=1}^t$ to learn the dictionaries D_{MR}, D_L, D_S can represent the corresponding samples using the common coefficients matrix $A^{(C)} = \{\boldsymbol{\alpha}_i^{(C)}\}_{i=1}^t$. Specifically, the complete term is defined by:

$$\begin{aligned} &E_{complete}(D_{MR}, D_L, D_S, A^{(C)}) = \\ &LCC\left(D_{MR}, \left\{\boldsymbol{\alpha}_i^{(C)}\right\}_{i=1}^t, \left\{\boldsymbol{x}_i^{MR}\right\}_{i=1}^t\right) \\ &+ LCC\left(D_L, \left\{\boldsymbol{\alpha}_i^{(C)}\right\}_{i=1}^t, \left\{\boldsymbol{x}_i^L\right\}_{i=1}^t\right) \\ &+ LCC\left(D_S, \left\{\boldsymbol{\alpha}_i^{(C)}\right\}_{i=1}^t, \left\{\boldsymbol{x}_i^S\right\}_{i=1}^t\right) \end{aligned} \qquad (9)$$

The incomplete terms are defined as follows:

$$E_{MR}\left(D_{MR}, A^{(MR)}\right) = LCC\left(D_{MR}, \left\{\boldsymbol{\alpha}_i^{(MR)}\right\}_{i=t+1}^{N_{MR}}, \left\{\boldsymbol{x}_i^{MR}\right\}_{i=t+1}^{N_{MR}}\right),$$
$$E_L\left(D_L, A^{(L)}\right) = LCC\left(D_L, \left\{\boldsymbol{\alpha}_i^{(L)}\right\}_{i=t+1}^{N_L}, \left\{\boldsymbol{x}_i^{L}\right\}_{i=t+1}^{N_L}\right),$$
$$E_S\left(D_S, A^{(S)}\right) = LCC\left(D_S, \left\{\boldsymbol{\alpha}_i^{(S)}\right\}_{i=t+1}^{N_S}, \left\{\boldsymbol{x}_i^{S}\right\}_{i=t+1}^{N_S}\right). \tag{10}$$

Although Eq. (8) is complex, it is convex over D when A is fixed and vice versa. Similar to LCC, Eq. (8) can be solved by dividing it into two sub-problems and optimizing D and A alternatively, as explained below.

Updating $A^{(C)}$, $A^{(MR)}$, $A^{(L)}$, $A^{(S)}$: First, we fix D_{MR}, D_L, and D_S to update the sparse coefficients $A^{(C)}$, $A^{(MR)}$, $A^{(L)}$, $A^{(S)}$. The dictionary initializations are performed in a similar manner as that in Section 3.1.1. The individual sparse coefficients $A^{(MR)}$, $A^{(L)}$ and $A^{(S)}$,

using three individual modality patch sets $\left\{\boldsymbol{x}_i^{MR}\right\}_{i=t+1}^{N_{MR}}, \left\{\boldsymbol{x}_i^{L}\right\}_{i=t+1}^{N_L}, \left\{\boldsymbol{x}_i^{S}\right\}_{i=t+1}^{N_S}$, can be directly obtained via the following three LASSO problems when fixing D_{MR}, D_L, and D_S. Here, for conciseness, we give one equation with a variable (Q) that can be replaced by MR, L or S,

$$\arg\min_{A^{(Q)}} LCC\left(D_Q, \left\{\boldsymbol{\alpha}_i^{(Q)}\right\}_{i=t+1}^{N_Q}, \left\{\boldsymbol{x}_i^{(Q)}\right\}_{i=t+1}^{N_Q}\right)$$
$$= \arg\min_{\boldsymbol{\beta}_i^{(Q)}} \sum_{i=t+1}^{N_Q} \left(\frac{1}{2}\left\|\boldsymbol{x}_i^{Q} - D_Q \Lambda_{Q,i}^{-1} \boldsymbol{\beta}_i^{(Q)}\right\|^2 + \mu\left\|\boldsymbol{\beta}_i^{(Q)}\right\|_1\right) \tag{11}$$

where $\Lambda_{Q,i}$ is a diagonal matrix whose diagonal elements are $\Lambda_{Q,i}^q = \left\|\boldsymbol{d}_Q^q - \boldsymbol{x}_i^Q\right\|^2$, $\boldsymbol{d}_Q^q$ is the q-th column of D_Q, $q = 1, \ldots, P$, and $\boldsymbol{\alpha}_i^{(Q)} = \Lambda_{Q,i}^{-1} \boldsymbol{\beta}_i^{(Q)}$.

For the shared coefficient $A^{(C)}$ using the tripled patches, we can concatenate the tripled patches $\left\{\left(\boldsymbol{x}_i^{MR}, \boldsymbol{x}_i^{L}, \boldsymbol{x}_i^{S}\right)\right\}_{i=1}^{t}$ and the corresponding dictionaries (D_{MR}, D_L and D_S) to jointly learn $A^{(C)}$, which can be obtained using the following objective function of LCC.

$$\arg\min_{A^{(C)}} LCC\left(\begin{bmatrix} D_{MR} \\ D_L \\ D_S \end{bmatrix}, \left\{\boldsymbol{\alpha}_i^{(C)}\right\}_{i=1}^{t}, \left\{\begin{bmatrix} \boldsymbol{x}_i^{MR} \\ \boldsymbol{x}_i^{L} \\ \boldsymbol{x}_i^{S} \end{bmatrix}\right\}_{i=1}^{t}\right)$$
$$= \arg\min_{\boldsymbol{\beta}_i^{(C)}} \sum_{i=1}^{t} \left(\frac{1}{2}\left\|\begin{bmatrix} \boldsymbol{x}_i^{MR} \\ \boldsymbol{x}_i^{L} \\ \boldsymbol{x}_i^{S} \end{bmatrix} - \begin{bmatrix} D_{MR} \\ D_L \\ D_S \end{bmatrix} \Lambda_{C,i}^{-1} \boldsymbol{\beta}_i^{(C)}\right\|^2 + \mu\left\|\boldsymbol{\beta}_i^{(C)}\right\|_1\right) \tag{12}$$

where $\Lambda_{C,i} = \Lambda_{MR,i} + \Lambda_{L,i} + \Lambda_{S,i}$ and $\boldsymbol{\alpha}_i^{(C)} = \Lambda_{C,i}^{-1} \boldsymbol{\beta}_i^{(C)}$.

Updating Dictionaries D$_{MR}$, D$_L$, D$_S$: Given the fixed sparse coefficients $A^{(C)}$, $A^{(MR)}$, $A^{(L)}$, $A^{(S)}$, the dictionaries D_{MR}, D_L and D_S can be optimized individually by connecting the complete samples and the incomplete samples in a coherent manner, and also connecting the corresponding fixed coefficients in a similar way. Therefore, D_{MR}, D_L, and D_S can be optimized as follows:

$$\underset{D_{MR}}{\arg\min} LCC\left(D_{MR}, \left\{\boldsymbol{\alpha}_i^{(MR)}\right\}_{i=1}^{N_{MR}}, \left\{\boldsymbol{x}_i^{MR}\right\}_{i=1}^{N_{MR}}\right)$$
$$\underset{D_L}{\arg\min} LCC\left(D_L, \left\{\boldsymbol{\alpha}_i^{(L)}\right\}_{i=1}^{N_L}, \left\{\boldsymbol{x}_i^{L}\right\}_{i=1}^{N_L}\right),$$
$$\underset{D_S}{\arg\min} LCC\left(D_S, \left\{\boldsymbol{\alpha}_i^{(S)}\right\}_{i=1}^{N_S}, \left\{\boldsymbol{x}_i^{S}\right\}_{i=1}^{N_S}\right). \tag{13}$$

In summary, by alternatively updating the dictionaries D_{MR}, D_L, D_S and the sparse coefficients $A^{(C)}$, $A^{(MR)}$, $A^{(L)}$, $A^{(S)}$, we can finally obtain three dictionaries D_{MR}, D_L and D_S. In summary, by alternatively updating the dictionaries D_{MR}, D_L, D_S the sparse coefficients $A^{(C)}$, $A^{(MR)}$, $A^{(L)}$, $A^{(S)}$, we can finally obtain three dictionaries D_{MR}, D_L and D_S.

For better understanding the tripled dictionary training procedure, we give a pseudo code, as shown in Algorithm 1.

Algorithm 1

Tripled Dictionary Training

1: **Input:** Training patch sets $\left\{\boldsymbol{x}_1^{MR}, \ldots, \boldsymbol{x}_i^{MR}, \ldots, \boldsymbol{x}_{N_{MR}}^{MR}\right\} \in \mathbb{R}^{H \times N_{MR}}$, $\left\{\boldsymbol{x}_1^{L}, \ldots, \boldsymbol{x}_i^{L}, \ldots, \boldsymbol{x}_{N_L}^{L}\right\} \in \mathbb{R}^{H \times N_L}$, and $\left\{\boldsymbol{x}_1^{S}, \ldots, \boldsymbol{x}_i^{S}, \ldots, \boldsymbol{x}_{N_S}^{S}\right\} \in \mathbb{R}^{H \times N_S}$, and dictionary size P

2:               **Initial:** Initialize $D_{MR}^{(0)}, D_L^{(0)}$ and $D_S^{(0)} \in \mathbb{R}^{H \times P}$.

3.   **Repeat**

4:               **For** $i = 1, 2, \ldots, t$ (t is the number of tripled patches) **do**

5:                   Compute $\boldsymbol{\alpha}_i^{(C)}$ according to Eq. (12);

6:               **End**

7:               **For** $i = t+1, t+2, \ldots, N_Q$ (Q can be replaced by MR, L or S) **do**

8:                   Compute $\boldsymbol{\alpha}_i^{(\mathscr{Q})}$ according to Eq. (11);

9:               **End**

10:              Update D_{MR}, D_L, and D_S, according to Eq. 13.

11:           **Until** convergence

12:           **Output:** Tripled dictionaries D_{MR}, D_L, and D_S.

---

2) SSTDL for S-PET Image Prediction—Given a testing sample with MR and L-PET images, for voxel z to be predicted, we respectively extract the patches $\boldsymbol{x}_{test}^{MR}$ and $\boldsymbol{x}_{test}^{L}$ from

the testing MR and L-PET images. Note that x_{test}^{MR} and x_{test}^{L} have been processed as described in Section 2. Then, we can jointly encode them by the LCC sparse representation using the learned dictionaries D_{MR} and D_L. Specifically, a sparse coefficient vector α_{test} is calculated as follows:

$$arg \min_{\alpha_{test}} LCC \left(\begin{bmatrix} D_{MR} \\ D_L \end{bmatrix}, \alpha_{test}, \begin{bmatrix} x_{test}^{MR} \\ x_{test}^{L} \end{bmatrix} \right)$$
$$= arg \min_{\beta_{test}} \frac{1}{2} \left\| \begin{bmatrix} x_{test}^{MR} \\ x_{test}^{L} \end{bmatrix} - \begin{bmatrix} D_{MR} \\ D_L \end{bmatrix} \Lambda_{test}^{-1} \beta_{test} \right\|^2 + \mu \| \beta_{test} \|_1 \quad (14)$$

where Λ_{test} is a diagonal matrix whose elements are $\Lambda_{test}^{q} = \left\| d_{MR}^{q} - x_{test}^{MR} \right\|^2 + \left\| d_{L}^{q} - x_{test}^{L} \right\|^2$, $q = 1, ..., P$, and $\alpha_{test} = \Lambda_{test}^{-1} \beta_{test}$.

Then, we apply the estimated sparse coefficient vector α_{test} on D_S to predict the S-PET image.

$$x_{test}^{S} = D_S \alpha_{test} \quad (15)$$

To determine the final prediction value at voxel z of the *unknown* S-PET image of the testing subject, one simple way is by taking the center value from the predicted patch x_{test}^{S}. To preliminary evaluate the proposed method, we use the center value for the final prediction in the Experiments 4.1, 4.2 and 4.3. However, in Ref. [24], it is suggested to average the results of the overlapping patches to determine the final value. To compare the performance of the two estimation methods, we further average the overlapping prediction values on each voxel to its final prediction in the final experiment (Section 4.4).

IV. Experimental Results

Dataset

As an exploration study, we tried our best and recruited 18 healthy adult volunteers for PET and MRI imaging scan. In this dataset, eight subjects have the complete data (i.e., having all L-PET, S-PET and MR images from these three modalities), and the other 10 subjects have the incomplete data (i.e., without L-PET images). The detailed demographic information of these subjects is summarized in Table I.

All scans were acquired by a Siemens Biograph mMR system housed in the Biomedical Research Imaging Center, and this study has been approved by the University of North Carolina at Chapel Hill (UNC) Institutional Review Board. As mentioned above, the standard-dose scans in our cohort was administered an average of 203 MBq (from 191 MBq to 229 MBq) of ^{18}F-2-deoxyglucose (^{18}FDG), which corresponds to an effective dose of 3.86 mSv and is in the low part of the range recommended by the Society of Nuclear Medicine and Molecular Imaging for FDG brain PET. The mean time post-injection was 36

minutes (from 32 to 41 minutes). S-PET and L-PET acquisitions were performed consecutively: first 12-minute S-PET immediately followed by 3-minute L-PET scans. The ordering was necessarily fixed due to the need to acquire clinical PET scans first and then experimental scans. This ordering creates a few implications. One of them is that the standard uptake value (SUV) increases slightly for the L-PET scans because the uptake in the brain continues to increase measurably during the acquisition time. Also, the effective noise level for the L-PET scans is slightly higher due to the radioactive decay that occurs during the acquisition time. We cannot say that these two effects cancel out, but they tend to work against each other to create L-PET datasets that are the equivalent of approximately one-quarter of the S-PET dose. In data acquisitions, no head holders were used. Subjects were simply asked to remain still. Resulting images were visually checked for apparent motion by examining the alignment of early S-PET and late L-PET images with the attenuation map, and only the cases that passed that check were used in the dataset. Both S-PET and L-PET images of the same subject used the same attenuation map, which was acquired prior to the S-PET scan. Note that the attenuation maps were computed from the Dixon fat-water method provided by the scanner manufacturer. Iterative reconstruction was employed with the ordered subsets expectation maximization (OSEM) algorithm [25], with three iterations, 21 subsets, and post-reconstruction filtered with a 3D Gaussian with FWHM of 2 mm. Meanwhile, T1-weighted images and diffusion-weighted images (DTI) were also acquired. For each subject, the PET images and the DTI images are, respectively, co-registered to the T1 image via affine transformation [26]. After image alignment, all images have the same dimensions of 128×128×128 and the same voxel size of 2.09×2.09×2.03 mm^3. Non-brain tissues were then removed from the aligned images using skull stripping [27], and the MR and L-PET images were preprocessed using intensity normalization to get more accurate sparse coefficients. A leave-one-out cross-validation (LOOCV) strategy is employed to evaluate the proposed method. In the testing stage, since the testing sample should have both MR and L-PET modalities, only the subjects with complete modalities can be used as the testing subject(s) in LOOCV.

Parameters

The selection of common parameters, such as patch size and neighborhood range, have been well studied (for details, see Ref. [28, 29]). Based on the suggested parameter settings as well as empirically validated in our experiments, the patch size and neighborhood range are set as 5×5×5 and 15×15×15, respectively. In the mapping procedure, the parameter γ with a value of 0.8 was found to be optimal [30]. As reported in Ref. [30], the parameter μ in SSTDL method for l_l-norm strength is set to 0.1, to balance the reconstruction error and local penalty. Taking into account both the computational efficiency and the convergence of SSTDL method, we set the number of iterations N_A for updating the atoms of dictionary to 500 and the total number of iterations N_D to 20. To ensure the redundancy of the learned dictionary in SSTDL method, we set the number of dictionary atoms to 512. The parameters for the SSTDL method used in the experiments are summarized in Table II.

Evaluations

To quantitatively evaluate the proposed method, three metrics were used for performance evaluation:

1) Normalized mean square error (NMSE): This is used to measure the voxel-wise intensity differences between the predicted S-PET image $\hat{I}^S$ with the ground truth I^S (i.e., the original S-PET image).

$$\text{NMSE} = \frac{\left\| I^S - \hat{I}^S \right\|_2^2}{\left\| I^S \right\|_2^2} \quad (16)$$

2) Peak signal-to-noise ratio (PSNR): This is used to evaluate the prediction accuracy in terms of the logarithmic decibel scale.

$$\text{PSNR} = 10 \log_{10} \left(\frac{R^2}{\frac{1}{U} \left\| I^S - \hat{I}^S \right\|_2^2} \right) \quad (17)$$

where R is the maximum intensity range of images I^S and $\hat{I}^S$, represents the total number of voxels in the image. Lower NMSE and higher PSNR may suggest a high-quality prediction.

3) Contrast recovery (CR) correlation: Contrast recovery (CR) is used to assess the contrast level between the ROI and background.

$$\text{CR} = \left(1 - \frac{\text{mean(ROI)}}{\text{mean(Background)}} \right) \quad (18)$$

where mean(ROI) represents the mean value of an ROI, and mean(Background) represents the mean value of the background. To compare the predicted S-PET and the ground-truth S-PET, we calculate CR correlation, which evaluates the relevance of CR between the predicted S-PET image and the ground-truth S-PET:

$$\text{CR correlation} = \frac{\sum_i \left(\hat{c}_i^S - \bar{\hat{c}}^S \right) \left(c_i^S - \bar{c}^S \right)}{\sqrt{\sum_i \left(\hat{c}_i^S - \bar{\hat{c}}^S \right)^2} \sqrt{\sum_i \left(c_i^S - \bar{c}^S \right)^2}}, \quad (19)$$

where $\hat{c}_i^S$ and c_i^S are respectively the CR of the predicted S-PET image and the ground-truth S-PET of the i-th subject, $\bar{\hat{c}}_i^S$ and $\bar{c}^S$ are respectively the average CR of the predicted S-PET images and the ground-truth S-PET images over all the subjects. In this paper, the hippocampus regions, which play an important role in the consolidation of information from short-term memory to long-term memory and spatial navigation [31], are used as the ROIs. And the cerebellum region was chosen as the background. To define the ROIs, we labeled the T1

images with 2 ROIs, hippocampus and cerebellum, by using a multi-atlas based segmentation method. Specifically, 15 MR brain images from the OASIS project and their corresponding label maps as provided by Neuromorphometrics, Inc. (http://Neuromorphometrics.com/) under academic subscription were selected as atlases. For a target image to be labeled, we registered multiple atlases onto the target image space (using FLIRT and Demons registration methods), and then used the estimated transformations to warp the corresponding label maps of atlases. Finally, we made use of a majority voting scheme on the warped label maps of all atlases, and then obtained the label map for the given target image. Ideally, the CR in the predicted S-PET image should be the same to the CR in the ground-truth S-PET; thus, the higher the CR correlation is, the better the prediction is.

A. S-PET Image Prediction using SSTDL

We first compare SSTDL with the SR method. Note that, for both methods, the data were processed according to the procedures described in Section 2, with both mapping and MRI fusion. For SR method, only the complete samples were used to construct the dictionaries, while, in the proposed SSTDL method, both complete and incomplete samples were used to jointly learn the dictionaries. Sample images of the prediction by SR and SSTDL are shown in Fig. 4.

As shown in Fig. 4, there is strong noise in the L-PET image. Compared to the S-PET image, some regions of the L-PET image are more fuzzy. This suggests that, with a reduced dose, the reconstructed L-PET has an inferior visibility compared to the desired S-PET images. The quality of the prediction by SR and SSTDL methods is substantially better based on visual observation than that of L-PET. Compared to SR, the prediction from our method is closer to the ground truth. For SR method, over-smoothing appears in some regions, which result in losing of some detailed information, as indicated by red arrows in Fig. 4. For quantitative comparison, we computed NMSE and PSNR between the predicted S-PET image and the ground-truth S-PET, and the results are given in Fig. 5. Table III lists the CR correlation scores.

From Fig. 5, we can see that, compared to the SR method, SSTDL consistently outperforms across different subjects, suggesting that the proposed method is more effective in predicting the S-PET image from MR and L-PET images. Meanwhile, the CR correlation of our method is higher than that of SR method as shown in Table 3, indicating that the prediction by SSTDL is much closer to the ground-truth S-PET. In summary, compared to the SR method, our method achieves better results due to the utilization of incomplete samples that are typically overlooked in the SR method.

B. Influence of Important Elements in SSTDL

We now investigate the influence of the key elements of the proposed method on prediction. Such elements include 1) *the use of multimodal MR images for prediction*, 2) *the mapping strategy*, and 3) *the use of incomplete-modality samples for prediction*.

1) Influence of the Use of Multimodal MR Images for Prediction—In this experiment, we use, respectively, (1) L-PET alone, (2) the combination of single modal MR and L-PET images (T1+L-PET), and (3) the combination of multimodal MR and L-PET images (T1+FA+MD+L-PET), for S-PET image prediction. The results in terms of PSNR and NMSE are given in Fig. 6.

As shown in Fig. 6, if using only L-PET image, the prediction has the lowest image quality, implying that the additional information provided by MR images is critical for improving the prediction performance. Although the improvement can be obtained by using single modal MR image (i.e., T1) together with the L-PET image, the results are still not as good as those obtained by joint use of all multimodal MR images and the L-PET image. These results indicate the importance of using multimodal MR images (if available) for improvement of S-PET image prediction.

2) Influence of the Mapping Procedure—To show the efficacy of the mapping procedure, we should compare the results of the SSTDL method *with* and *without* the mapping procedure. However, the multimodal MR fusion would no longer make sense without the mapping procedure. Therefore, to make the comparison more meaningful, we just consider the L-PET images for prediction in this experiment. Specifically, in the first case, the L-PET images are first mapped and then used for prediction of S-PET image based on the proposed method. In the second case, the L-PET images are directly used for prediction of S-PET image without the mapping procedure. The comparison results in terms of PSNR and NMSE are shown in Fig. 7.

Fig. 7 shows that, compared with the results without mapping, our results with mapping achieve higher PSNR and lower NMSE across all subjects. This clearly indicates that the mapping strategy is necessary and can significantly improve the prediction accuracy.

3) Influence of the Use of Incomplete-modality Samples for S-PET Image Prediction—One of the greatest advantages of the proposed SSTDL method is that, besides the complete-modality samples, it can also leverage the incomplete-modality samples for prediction. Therefore, we study the prediction performance of using the incomplete samples. Specifically, we conduct an experiment to *compare* the performance of SSTDL method using only the complete samples *with* that using both the complete and the incomplete samples. Fig. 8 shows the influence of the incomplete samples in S-PET image prediction, and the results suggest that the incomplete samples benefit the prediction and should be utilized if available.

Table IV compares the CR correlation using different key strategies in the proposed method.

From Table IV, we can see that the CR correlation of SSTDL using 1) mapping strategy, 2) MR images, and 3) incomplete-modality samples is the highest, which is consistent with the observations as aforementioned.

C. Compared with learning-based method: Random Forest (RF)

In Ref. [32], the random forest method has been successfully used in S-PET prediction. Random Forest (RF), which was originally proposed by Breiman [33], is an ensemble learning method for classification or regression that operates by constructing a multitude of binary decision trees. Each tree is trained independently with random features and thresholds, and this ensemble produces a corresponding number of outputs. Then, one final prediction is obtained by aggregating the outputs of all trees. In the following experiment, we compare the performance of RF with our proposed method. The same processing scheme is performed for random forest, i.e., we also use the combination of MR and L-PET images to predict the S-PET image. The parameters of RF are set as follows: the patch size: 5×5×5; the neighborhood size: 15×15×15; the number of trees in a forest: 10; the number of randomly selected features: 1000; the maximum tree depth: 15; and the minimum number of samples at each leaf: 5. The comparison results are shown in Fig. 9.

As shown in Fig. 9, although the results of subject 1 of RF is a little better than our proposed method (with higher PSRN and lower NMSE), our proposed method generally achieves a more accurate predictions than RF, as the average PSNR increases by approximately 1.4. Therefore, by comparing with this learning based method, we further demonstrate the effectiveness of our proposed method.

D. Comparison with the overlapped average estimation for final S-PET prediction

For the above experiments, the central value of the predicted patch was used for the final prediction. In Ref. [42], however, it was suggested to average the results of the overlapping patches for determining the final prediction. In this section, we further evaluate the performance of using the overlapped average value as discussed in the last paragraph of Section 3.2.2, with the results given in Fig. 10.

From Fig. 10, we can observe that, for four subjects (subject 1, 4, 5 and 8), the results using the average value of the overlapping patches outperform the method using the central value as the final prediction. Besides, for the other four subjects, using the central value indicates a better performance. Then, for general evaluation, the averaged PSNR and NMSE across the eight subjects are calculated. The results indicate that, using the overlapped average estimation, the average PSNR increases by 0.2 while the average NMSE decrease by 0.0003, suggesting that the overlapped average estimation method could be a more effective method for the final prediction.

V. Conclusion

In this paper, we have proposed a semi-supervised tripled dictionary learning (SSTDL) method for predicting the S-PET image from multimodal MR and L-PET images. On a real dataset of human brain images, we performed rigorous experiments, and both quantitative and qualitative results suggested that the proposed method outperforms the baseline methods using partial data. The predicted S-PET images by our method are very close to the ground-truth S-PET. The obtained experimental results demonstrate its great potential in real-world clinical applications.

It is worth noting that, different from existing techniques that aim to improve the PET image quality during the reconstruction from raw signals, the proposed method aims to directly estimate the S-PET image from its low-dose counterpart and corresponding MR image. This approach has rarely been studied previously in the literature. The proposed method can utilize not only the samples with complete data but also the samples with incomplete data. This property can take advantage of the large number of available training samples, which cannot be handled by other techniques such as sparse representation and random forest.

However, all the brain scans used in this paper are the normal subjects, and the PET and MR images are well correlated. Thus, at this stage we cannot comment on how our method might perform on regions other than brain, patients with tumors, or for other applications, body locations, PET tracers, dose-reduction ratios, or scanners. As a result, our future work will focus on evaluating the proposed method on the scans with more variations (e.g., different ages, abnormal anatomy due to tumors, atrophy and locally resected brain). In addition, the proposed method is computationally expensive since complex optimization process is required, we are now in the progress of working for a more efficient optimization method.

Acknowledgments

References

1. Kitson SL, Cuccurullo V, Ciarmiello A, Salvo D, Mansi L. Clinical applications of positron emission tomography (PET) imaging in medicine: oncology, brain diseases and cardiology. Current Radiopharmaceuticals. Oct; 2009 2(4):224–253.

2. Sebro R, Mari-Aparici M, Mari-Aparici C. Value of true whole-body FDG-PET/CT scanning protocol in oncology: Optimization of its use based on primary diagnosis. Acta Radiol. Jun; 2013 54(5):534–539. [PubMed: 23463863]

3. Sarikaya I. Cardiac applications of PET. Nucl Med Commun. Oct; 2015 36(10):971–985. [PubMed: 26035516]

4. Casse R, Rowe CC, Newton M, Berlangieri SU, Scott AM. Positron emission tomography and epilepsy. Mol Imaging Biol. Oct; 2002 4(5):338–351. [PubMed: 14537108]

5. Chen W. Clinical applications of PET in brain tumors. J Nucl Med. Aug; 2007 48(9):1468–1481. [PubMed: 17704239]

6. Quigley H, Colloby SJ, O'Brien JT. PET imaging of brain amyloid in dementia: a review. International Journal of Geriatric Psychiatry. Oct; 2011 26(10):991–999. [PubMed: 21905095]

7. Bai B, Li Q, Leahy RM. Magnetic resonance-guided positron emission tomography image reconstruction. Seminars in Nuclear Medicine. Jan; 2013 43(1):30–44. [PubMed: 23178087]

8. Yang J, Wright J, Huang TS, et al. Image super-resolution via sparse representation. IEEE Trans Image Process. May; 2010 19(11):2861–2873. [PubMed: 20483687]

9. Elad M, Aharon M. Image denoising via learned dictionaries and sparse representation. IEEE Comput Soc Conf Comput Vis Pattern Recognit. Jun.2006 1:895–900.

10. Elad M, Aharon M. Image denoising via sparse and redundant representations over learned dictionaries. IEEE Trans Image Process. Dec; 2006 15(12):3736–3745. [PubMed: 17153947]

11. Wright J, Yang AY, Ganesh A, Sastry SS, Ma Y. Robust face recognition via sparse representation. IEEE Trans Pattern Anal Mach Intell. Feb; 2009 31(2):210–227. [PubMed: 19110489]

12. Valiollahzadeh S, Clark J, Mawlawi O. PET Image Deblurring using Adaptive Dictionary Learning. Med Phys. Jan.2014 41(1)

13. Chen S, Liu H, Shi P, et al. Sparse representation and dictionary learning penalized image reconstruction for positron emission tomography. Phys Med Biol. Jan; 2015 60(2):807–823. [PubMed: 25565039]

14. Lu Y, Zhao J, Wang G. Few-view image reconstruction with dual dictionaries. Phys Med Biol. Jan; 2012 57(1):173–189. [PubMed: 22155989]

15. Lu Y, Zhao J, Zhuang T, Wang G. Unified dual-modality image reconstruction with dual dictionaries. Proc SPIE. 2012

16. Zhao B, Ding H, Lu Y, Wang G, Zhao J, Molloi S. Dual-dictionary learning-based iterative image reconstruction for spectral computed tomography application. Phys Med Biol. Nov; 2012 57(24): 8217–8229. [PubMed: 23192234]

17. Valiollahzadeh S, Clark JW, Mawlawi O. Dictionary learning for data recovery in positron emission tomography. Phys Med Biol. Aug; 2015 60(15):5853–5871. [PubMed: 26161630]

18. Yu K, Zhang T, Gong Y. Nonlinear learning using local coordinate coding. Advances in NIPS. 2009:2223–2231.

19. Wang D, Steven C, He Y, Zhu J, Mei T, Luo J. Retrieval-based face annotation by weak label regularized local coordinate coding. IEEE Trans Pattern Anal Mach Intell. Mar; 2014 36(3):550–563. [PubMed: 24457510]

20. Zhou X, Yu K, Zhang T, Huang TS. Image classification using super-vector coding of local image descriptors. Computer Vision-ECCV. 2010:141–154.

21. Song M, Tao D, Sun S, Chen C, Maybank SJ. Robust 3D Face Landmark Localization based on Local Coordinate Coding. IEEE Trans Image Process. Dec; 2014 23(12):5108–5122. [PubMed: 25296404]

22. Xie B, Song M, Tao DC. Large-scale Dictionary Learning For Local Coordinate Coding. Proc BMVC. 2010:1–10.

23. Tibshirani R. Regression shrinkage and selection via the lasso. J Roy Stat Soc (Ser B). 1996:267–288.

24. Elad M, Aharon M. Image denoising via sparse and redundant representations over learned dictionaries. IEEE Trans Image Process. Dec; 2006 15(12):3736–3745. [PubMed: 17153947]

25. Hudson HM, Larkin RS. Accelerated image reconstruction using ordered subsets of projection data. IEEE Trans Med Imag. Dec; 1994 13(4):601–609.

26. Smith SM, Jenkinson M, Woolrich MW, Beckmann CF, Behrens TE, Johansen-Berg H, et al. Advances in functional and structural MR image analysis and implementation as FSL. Neuroimage. 2004:S208–S219. [PubMed: 15501092]

27. Shi F, Wang L, Dai Y, Gilmore JH, Lin W, Shen D. LABEL: Pediatric brain extraction using learning-based meta-algorithm. Neuroimage. Sep; 2012 62(3):1975–1986. [PubMed: 22634859]

28. Rousseau F, Habas PA, Studholme C. A supervised patch-based approach for human brain labeling. IEEE Trans Med Imag. May; 2011 30(10):1852–1862.

29. Tong, T., Wolz, R., Hajnal, JV., Rueckert, D. MICCAI Workshop on Sparsity Techniques in Medical Imaging. 2012. Segmentation of Brain Images via Sparse Patch Representation.

30. Wang Y, Zhang P, An L, Ma G, Kang J, Wu X, Zhou J, Lalush DS, Lin W, Shen D. Predicting Standard-Dose PET Image from Low-Dose PET and Multimodal MR Images Using Mapping-Based Sparse Representation. Machine Learning in Medical Imaging. Oct.2015 9352:127–135.

31. Sultana R, Boyd-Kimball D, Poon HF, Cai J, Pierce WM, Klein JB, Merchant M, Markesbery WR, Butterfield DA. Redox proteomics identification of oxidized proteins in Alzheimer's disease hippocampus and cerebellum: an approach to understand pathological and biochemical alterations in AD. Neurobiol Aging. Nov; 2006 27(11):1564–1576. [PubMed: 16271804]

32. Kang J, Gao Y, Wu Y, Ma G, Shi F, Lin W, Shen D. Prediction of Standard-Dose PET Image by Low-Dose PET and MRI Images. Machine Learning in Medical Imaging. 2014; 8679:280–288.

33. Breiman L. Random forests. Machine Learning. Oct; 2001 45(1):5–32.

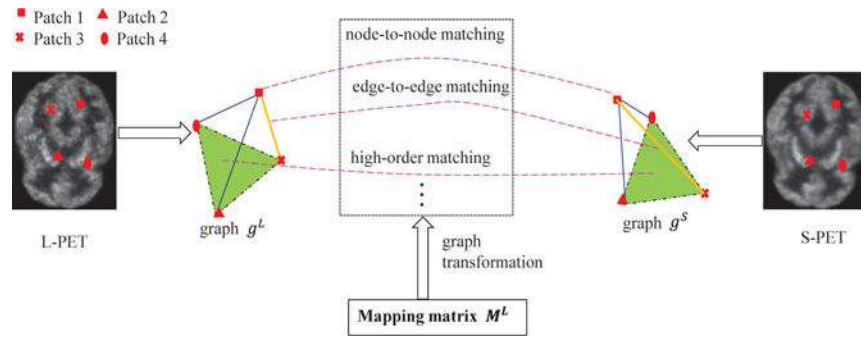**Fig. 1.**
The illustration of the mapping procedure between the L-PET and S-PET image patches.
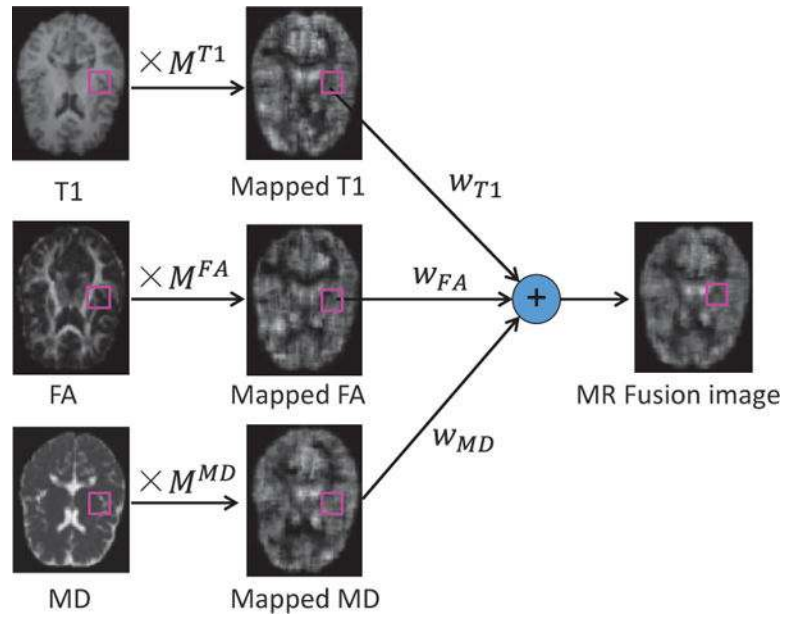
Fig. 2.
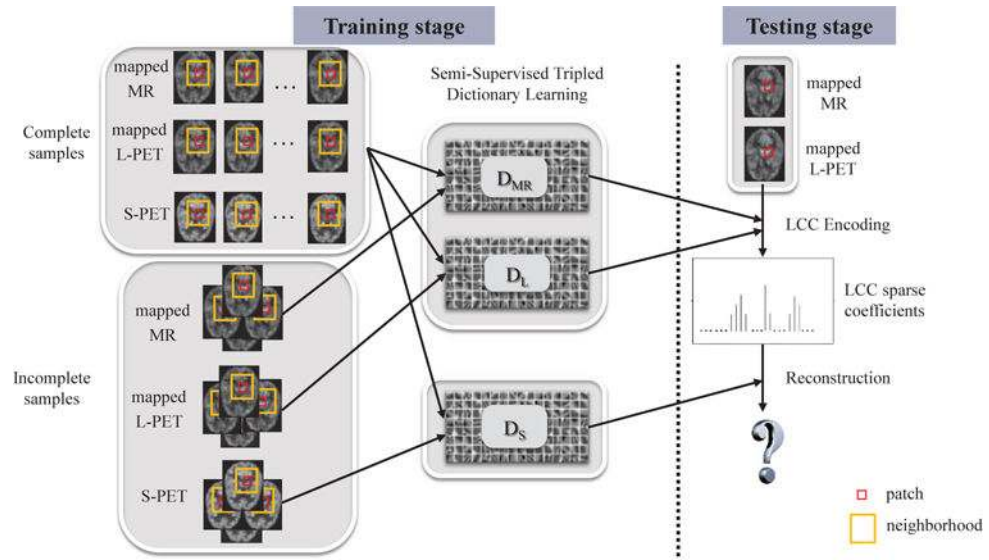Mapping and fusion procedures of the multimodal MR images.

Fig. 3.
Flowchart of the proposed semi-supervised tripled dictionary learning (SSTDL) method for S-PET prediction.
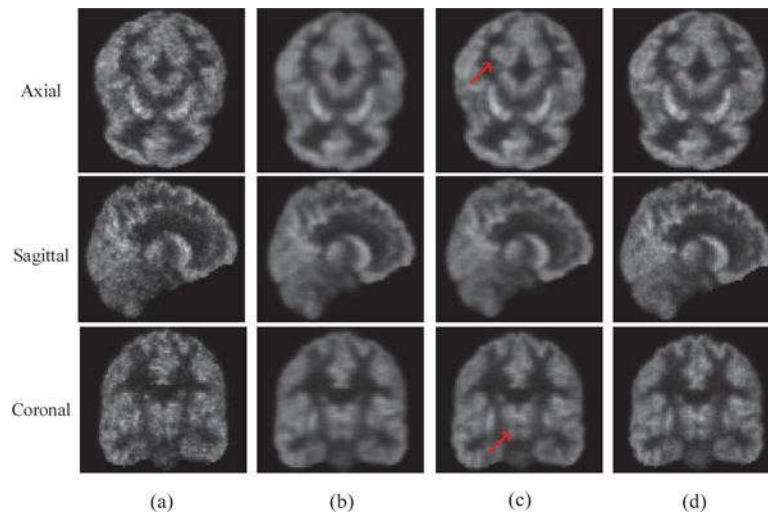
Fig. 4.
Comparison between the predicted S-PET images by the SR method and the proposed method: (a) L-PET, (b) SR, (c) SSTDL, and (d) ground truth.
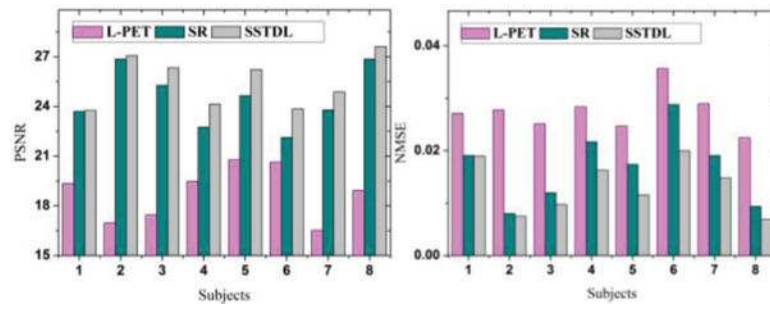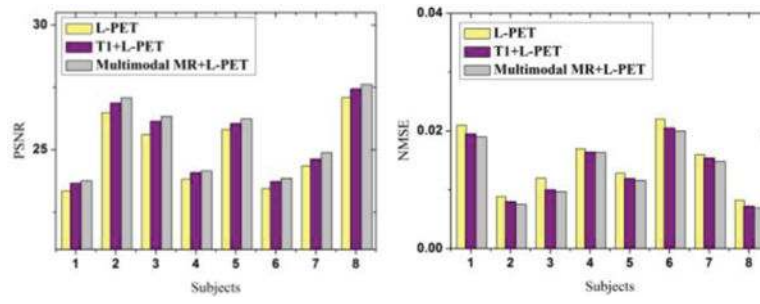
Fig. 5.
Comparison between SR and SSTDL in terms of PSNR and NMSE.

Fig. 6.
Influence of the use of multimodal MR images on the S-PET image prediction.
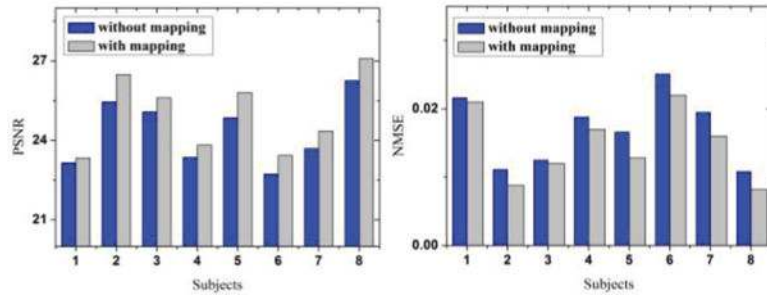
Fig. 7.
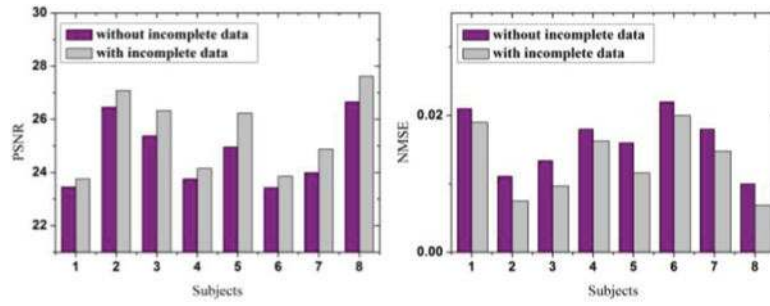Influence of the mapping procedure on the S-PET image prediction.

Fig. 8.
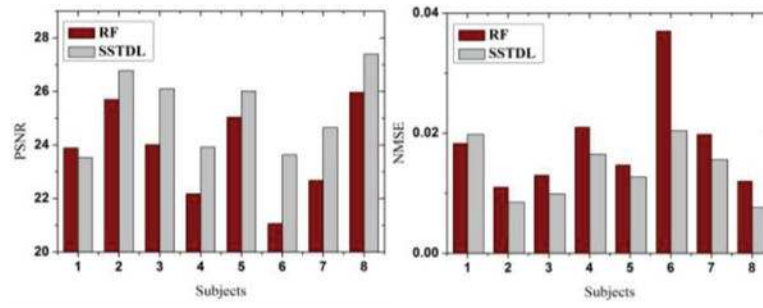Influence of the use of incomplete-modality samples on S-PET image prediction.

Fig. 9.
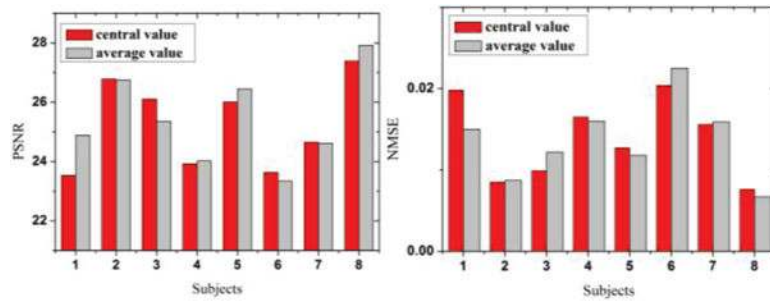Comparison between RF method and our proposed method.

Fig. 10.
Comparison with the overlapped average estimation for final S-PET prediction.

TABLE I

Demographic Information of The Subjects

	Complete samples	**Incomplete samples**
Total	8	10
Gender (Female/Male)	5/3	9/1
Age (Mean±SD)	26.38±6.16	26.2±7.13
Modality	MR, L-PET, S-PET	MR, S-PET

TABLE II

Summary of The Parameter Setting in The Experiment

Variable	Definition	Value
$\rho \times \rho \times \rho$	Patch size	5×5×5
$\omega \times \omega \times \omega$	Neighborhood size	15×15×15
γ	Parameter to balance the node-to-node matching and edge-to-edge matching in mapping procedure	0.8
μ	Parameter to balance the reconstruction error and locality penalty	0.1
P	Number of dictionary atoms	512
N_A	Number of iterations for updating the atoms of dictionary	500
N_D	Number of iterations for alternating optimization used in the dictionary learning	20

TABLE III

Comparison of CR Correlation

Method	L-PET	SR	SSTDL
CR correlation	0.9145	0.9369	**0.9520**

TABLE IV

CR Correlation Comparison of The Key Strategies in SSTDL

Method	SSTDL without mapping	SSTDL without MR images	SSTDL without incomplete samples	SSTDL (proposed)
CR correlation	0.9410	0.9486	0.9425	**0.9520**