IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received February 16, 2020, accepted March 26, 2020, date of publication April 1, 2020, date of current version April 15, 2020.

Digital Object Identifier 10.1109/ACCESS.2020.2984804

Unsupervised 3D PET-CT Image Registration Method Using a Metabolic Constraint Function and a Multi-Domain Similarity Measure

HENGJIAN YU[1], HUIYAN JIANG[1,2], XIANGRONG ZHOU[3], TAKESHI HARA[3], YU-DONG YAO[4], (Fellow, IEEE), AND HIROSHI FUJITA[3]

[1]Software College, Northeastern University, Shenyang 110819, China
[2]Key Laboratory of Intelligent Computing in Medical Image, Ministry of Education, Northeastern University, Shenyang 110819, China
[3]Department of Electrical, Electronic and Computer Engineering, Faculty of Engineering, Gifu University, Gifu 501-1193, Japan
[4]Department of Electrical and Computer Engineering, Stevens Institute of Technology, Hoboken, NJ 07030, USA

Corresponding author: Huiyan Jiang (hyjiang@mail.neu.edu.cn)

This work was supported by the National Natural Science Foundation of China under Grant 61872075.

ABSTRACT High-resolution CT images can clearly display anatomical structures but does not display functional information, while blurred PET images can display molecular and functional information of lesions but cannot clearly display morphological structures. Therefore, accurate PET-CT image registration, which is used for anatomical structure and functional information fusion, is a prerequisite for early stage cancer diagnosis. However, some hypermetabolic anatomical structures, such as brain and bladder, have low registration accuracy. To solve this problem, a 3D unsupervised network based on a metabolic constraint function and a multi-domain similarity measure (3D MC-MDS Net) is proposed for 3D PET-CT image registration. Specifically, a metabolic constraint model is established based on the standard uptake value (SUV) distribution of hypermetabolic regions such as brain, bladder, liver and heart, which reduces the excessive distortion on displacement vector field (DVF) caused by hypermetabolic anatomical structures in PET images. A DVF estimator is built based on 3D unsupervised convolutional neural networks and a spatial transformer is used for warping 3D PET images to 3D CT images. The generated registration results (PET image patches) and the original 3D CT image patches are used for calculating the spatial domain similarity (SD similarity) and frequency domain similarity (FD similarity). Finally, the loss function of the entire registration network is constructed by a weighted sum of SD similarity, FD similarity and a smoothness of DVF. A dataset consisted of 170 whole-body PET-CT images is used for registration accuracy evaluation. The proposed unsupervised registration network, 3D MC-MDS Net, can accurately learn the 3D registration model by using the training dataset with the metabolic constraint model, which significantly improves the registration accuracy.

INDEX TERMS PET/CT images, 3D image registration, unsupervised registration, metabolic constraint, convolutional neural network.

I. INTRODUCTION

Positron emission tomography (PET) scanners can track radioactive tracers to record the metabolic activity of patient anatomical structures, which is crucial for the detection and diagnosis of early stage cancer [1]. Because PET images are blurred and have low resolution, the integration of computed tomography (CT) scanners can further obtain the morphological information of patients. In general, the metabolic pattern of cancer tissues is more active than normal tissues. Therefore, compared with conventional imaging technology, the functional information and morphological information provided by PET-CT imaging is expected to significantly improve the sensitivity and specificity of early stage cancer diagnosis.

However, the present PET-CT image registration solution provided by vendors has decent accuracy in normal metabolic regions such as bone, but not ideal in the diaphragm and some

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

This work is licensed under a Creative Commons Attribution 4.0 License. For more information, see https://creativecommons.org/licenses/by/4.0/

hypermetabolic regions. Therefore, accurate PET-CT image registration is an important prerequisite for improving the accuracy of early stage cancer diagnosis.

To compensate for the anatomical structure mismatch in PET-CT caused by body movements and different postures, [2] proposed a cubic B-spline interpolation algorithm to capture local non-rigid movements of chest PET-CT images and mutual information (MI) was used to measure the similarity between current registration results and fixed images in an iterative process. Reference [3] performed a rigid registration on chest PET-CT images and evaluated the registration performance by measuring the average movement of anatomical structures. Reference [4] used a pyramid scheme combing global rigid registration and local non-rigid registration to perform a 3D registration of whole-body PET-CT images, in which the normalized mutual information (NMI) between fixed images (CT) and moving images (PET) was used as the similarity measure. Recently, [5] proposed an accurate registration framework for PET-CT and ultrasound volumes by concatenating a thin-plate spline based deformable registration method and a rigid registration algorithm, which was performed by maximizing the overlap between segmented prostate volume masks [6]. In these proposed methods, the range of registration is limited in local regions. Note that the method in [4] is 3D whole-body registration but is based on local non-rigid registration. In addition, the combination of rigid and non-rigid registration methods in a traditional iterative process increases the framework complexity and reduces the computational efficiency.

Mutual information can be used as a similarity measure in multimodal registration algorithms [7]–[10]. The mutual information value between images can be measured by considering the distribution of image gray values as one-dimensional signals. High mutual information values usually correspond to a high registration accuracy [11], but it cannot guarantee the complete alignment of local anatomical structures in whole-body PET-CT image registration tasks. The mutual information value between moving images (PET) and fixed images (CT) usually increases during the registration process. However, unlike other multimodal registration tasks, the non-shared information between PET images and CT images leads to an extra reduction in mutual information values, reflecting the difference between functional information in PET and anatomical structure information in CT. Because the anatomical structure corresponding to bright pixels in PET images can be over-deformed in non-rigid registration process, the original non-shared information between PET images and CT images is expected to be incorrectly transformed to shared information, resulting in further increase of the mutual information value. As a result, the ideal alignment does not correspond to the highest mutual information value in the above situation. Therefore, mutual information measures cannot be directly applied to whole-body PET-CT image registration tasks.

Recently, deep learning techniques [12] combined with iterative-based methods are used to obtain registration results. Reference [13] proposed a rigid registration model for MR images by using a convolutional neural network (CNN) as a similarity metric and the registration results outperformed MI based registration. Reference [14] proposed a supervised rigid image registration framework for 3D CT and cone-beam CT images and this end-to-end solution demonstrated the feasibility of deep learning technology in rigid registration tasks by using the output of CNN.

Iterative-based methods can obtain high registration accuracy but the registration efficiency is limited. Reference [15] proposed a real-time rigid image registration model by using CNN to generate transformation parameters and the registration results demonstrated that the proposed framework significantly improved the computational efficiency without losing much precision compared to conventional methods. Although supervised training models can be used in rigid image registration tasks, it is hard to obtain ground truth transformations corresponding to the training dataset. To overcome the difficulty of obtaining ground truth transformations, [16] used a stacked auto-encoder (SAE) for brain MR image registration and the proposed unsupervised approach avoided the tedious labeling process with human errors. However, the network training process is not end-to-end and limited in brain regions.

The reliance on ground truth transformations or parameters makes unsupervised registration an option, in which a key innovation called spatial transformer network (STN) [17] has been widely used [18]. STN can be used for fast feature map transformation and the differentiable characteristic makes it possible to be embedded in larger networks as a component. Reference [19] proposed a CNN based real-time deformable registration framework and the transformation parameters were used in STN to correct inspiration caused deformation in MR images. Inspired by STN, [20] concatenated a CNN regressor with a spatial transformer and proposed a 2D end-to-end deformable image registration network (2D DIR Net). 2D DIR Net expects a moving image and a fixed image as inputs and uses an unsupervised similarity measure as loss function for network training. The trained CNN regressor can automatically generate registration parameters with higher computational efficiency, avoiding the usage of ground truth transformation. However, 2D DIR Net cannot effectively analyze the information in 3D images for registration.

In order to accomplish automatic unsupervised registration on 3D PET-CT images, in our previous work, two registration methods (3D CNN-RT [22] and DenseRegNet [30]) were proposed based on deep learning technology. 3D CNN-RT utilized a 3D CNN network structure to register 3D PET-CT images and a regular term was added to the loss function for preventing the over-deformation of anatomical structures. DenseRegNet utilized a DenseNet based network structure for displacement vector field (DVF) predicting and proposed a two-level similarity measure for network training optimization, further improving the registration accuracy. Although, our previous works [22] [30] showed potential capability of a deep CNN model for image registration between PET and CT images, the accuracy of the registration on hypermetabolic regions such as the brain, bladder, liver and heart were still unsatisfied.

To solve the above problem, an unsupervised network is proposed in this paper for 3D PET-CT image registration by using a metabolic constraint function and a multi-domain similarity measure (3D MC-MDS Net). The main contributions of the proposed work are as follows.

- (1) A metabolic constraint model is built to restrict the over-deformation of hypermetabolic anatomical structures during the registration processes (Section II-A).
- (2) A novel frequency domain similarity (FD similarity) measure is combined with a spatial domain similarity (SD similarity) measure and a smoothness to construct loss function for the proposed registration network, which significantly improves the PET-CT image registration accuracy (Section II-C).

The remaining of this paper is organized as follows. Section II will describe the implementation of the proposed 3D MC-MDS Net. Section III will show the dataset and experimental process. Section IV will show the experimental results and discuss related issues. Section V will draw conclusions.

II. METHOD

In theory, the proposed method can perform registration from PET images to CT images, in which PET images are referred as moving images, and vice versa. However, PET images are blurred, which contain metabolic information. If the blurred PET images are referred as fixed images and the CT images are referred as moving images, it will lead to larger registration errors. Thus, the scheme of PET for moving images and CT for fixed images is selected.

A. METABOLIC CONSTRAINT MODEL

Some hypermetabolic anatomical structures in PET images cause the distortion in the PET-CT image registration process and reduce the registration accuracy [22], as shown in Figure 1.

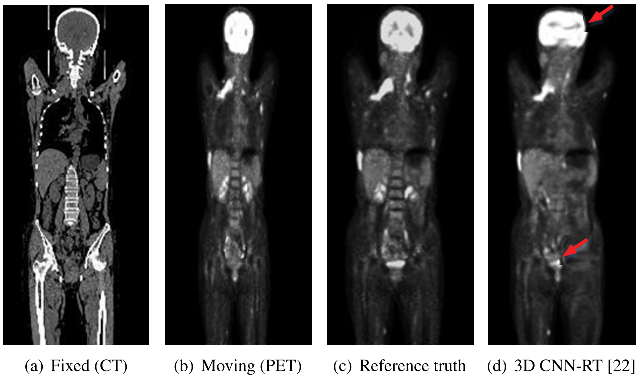

(a) Fixed (CT) (b) Moving (PET) (c) Reference truth (d) 3D CNN-RT [22]

FIGURE 1. The distortion caused by hypermetabolic anatomical structures (coronal section). (a) Fixed image (CT). (b) Moving image (PET). (c) Reference truth [28]. (d) 3D CNN-RT [22] registration result, the regions pointed by red arrows are distorted hypermetabolic anatomical structures (brain, bladder).

The registration results of hypermetabolic anatomical structures in Figure 1, which are pointed by arrows (brain and bladder), are severely distorted. The registration result in the brain region are excessively stretched in horizontal direction, causing severe distortion of the contour, while the registration result in the bladder region are significantly squeezed (Figure 1(d)). Therefore, reducing the adverse effects of hypermetabolic anatomical structure is needed for improving the registration accuracy.

To discover the standard uptake value (SUV) distribution pattern of hypermetabolic anatomical structures such as brain, bladder, heart and liver, a dataset consisted of 170 PET-CT images is used for statistical analysis. The SUV distribution in hypermetabolic anatomical structures is shown in Table 1 and Figure 2.

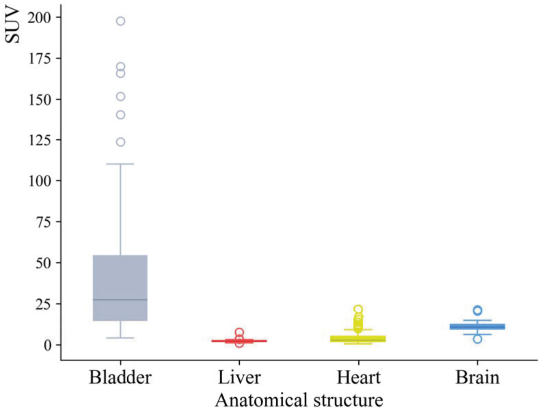


FIGURE 2. SUV distribution box diagram of hypermetabolic anatomical structures.

The bladder records the highest mean metabolic value ($\mu_{SUV} = 39.241$) in Table 1 and Figure 2, which is significantly higher than the metabolic level of the brain ($\mu_{SUV} = 11.056$) and the heart ($\mu_{SUV} = 3.970$). It can be explained by the pseudo-metabolic phenomenon generated by the deposition of the radioactive tracer, which contributes to the high bladder SUV and does not indicate the exact metabolic level. The standard deviation of bladder SUV is 34.772, which is much higher than brain ($\sigma_{SUV} = 3.185$) and heart ($\sigma_{SUV} = 3.346$), indicating that the deposition of the radiotracer amplifies the metabolic level differences between different individuals. As for liver, the mean and standard deviation of liver SUV are 2.144 and 0.604, respectively. Therefore, the registration distortion is severe in hypermetabolic anatomical structures with higher metabolic level such as brain and bladder in Figure 1(d), while the registration accuracy in anatomical structures with lower metabolic level is decent.

To suppress the registration error in hypermetabolic anatomical structures, a metabolic constraint model is proposed in this paper and applied to the registration process.

TABLE 1. SUV distribution of hypermetabolic anatomical structures.

Anatomical structure	Bladder	Liver	Heart	Brain
$\mu_{SUV}\pm\sigma_{SUV}$	39.241±34.772	2.144±0.604	3.970±3.346	11.056±3.185

The metabolic constraint function is as follows.

$$f_{mc}(p) = \begin{cases} k_1 \times p(x,y,z) + b_1, & if\ 0 \le p(x,y,z) \le \mu_1 \\ k_2 \times p(x,y,z) + b_2, & if\ \mu_1 \le p(x,y,z) \le \mu_2 \\ k_3 \times p(x,y,z) + b_3, & if,\ p(x,y,z) \ge \mu_2 \end{cases} \quad (1)$$

where f_{mc} represents the metabolic constraint function, $p(x, y, z)$ is the SUV in the moving image (PET), and x, y, z represent coordinates in the 3D space. k_i, b_i $(i = 1, 2, 3)$ and μ_j $(j = 1, 2)$ are hyperparameters.

To maintain the relationship between anatomical structures at different metabolic levels, the metabolic constraint function f_{mc} is monotonically increasing. Specifically, for $\forall p(x, y, z), q(x, y, z) \in M$, such that $p(x, y, z) \le q(x, y, z)$, then $f_{mc}(p(x, y, z)) \le f_{mc}(q(x, y, z))$, where M denotes a moving image.

B. 3D UNSUPERVISED REGISTRATION NETWORK STRUCTURE

The network structure of 3D MC-MDS Net is shown in Figure 3 and includes the following parts. (1) A metabolic constraint model (Section II-A), which is used to restrict the over-deformation of hypermetabolic anatomical structures during the registration processes; (2) A CNN based low-resolution displacement vector field (LR-DVF) estimator [22], which expects PET-CT images as inputs and outputs LR-DVF as transformation parameters; (3) A bicubic interpolation operation layer, which generates high-resolution displacement vector field (HR-DVF) from LR-DVF; (4) A 3D spatial transformer [23], which uses the HR-DVF to warp a moving PET image to a fixed CT image; (5) The multi-domain similarity loss for network training, which will be discussed in Section II-C.

The LR-DVF estimator is the only trainable structure in the entire network, which consists of 7 convolutional layers. The filter size of first 4 layers is 3 × 3 × 3@16 with an exponential linear unit (ELU) appended as activation function and the filter size of subsequent 2 convolutional layers is 1 × 1 × 1@16 with ELU appended. The size of the last convolutional layer is 1 × 1 × 1@3 without ELU function. In addition, each of the first 4 convolutional layers is linked to an average pooling layer of size 2 × 2 × 2.

To obtain accurate registration results, the resolution of HR-DVF in Figure 3 is set to be the same as original input patches. Since HR-DVF is directly obtained from LR-DVF by using bicubic interpolation, LR-DVF determines the deformation granularity of the moving image (PET) in the registration process. Lower the resolution of LR-DVF results in larger the deformation granularity and more components of rigid deformation in the registration process. On the contrary, high LR-DVF resolution corresponds to more non-rigid deformation components in the registration process. To make the balance between rigid deformation and non-rigid deformation, the resolution of LR-DVF is set to 4 × 4 × 4 in the training stage, which means that the resolution of LR-DVF in both horizontal and vertical directions is $1/16$ of HR-DVF.

The inputs of 3D MC-MDS Net in the training process are PET-CT image patches of size 64 × 64 × 64 and the dataset generation method is described in Section III-A. The network training process consists of the following steps.

(1) Generate 3D PET-CT image patches from whole-body PET-CT images by using overlapped sampling;

(2) Apply metabolic constraint model to 3D PET image patches by using metabolic constraint function (Eq. (1));

(3) Feed the normalized 3D PET image patches and 3D CT image patches jointly into the LR-DVF estimator, and calculating the smoothness of LR-DVF;

(4) Perform a bicubic interpolation on LR-DVF to generate the HR-DVF;

(5) Feed the HR-DVF and the original 3D PET patches jointly into the spatial transformer for 3D patch-level PET-CT registration;

(6) Calculate the spatial domain similarity (SD similarity) with the original 3D CT patches based on the current registration result;

(7) Perform a discrete wavelet transform (DWT) on the current registration result to generate 3 high frequency image patches and 1 low frequency image patch;

(8) Calculate the frequency domain similarity (FD similarity) between current registration results and the original 3D CT image patches;

(9) Construct the loss function of the entire registration network by weighting sum of the smoothness, the spatial domain similarity (SD similarity) and the frequency domain similarity (FD similarity);

(10) Perform back-propagation algorithm for LR-DVF estimator training.

The testing process of 3D MC-MDS Net is shown in the lower part of Figure 3. The input of the network are whole-body PET-CT images of size $h \times w \times d$, where h and w are height and width of the image in horizontal direction, d is depth of the image in vertical direction. The network testing process consists of the following steps.

(1) Apply metabolic constraint model to 3D whole-body PET images by using metabolic constraint function (Eq. (1));

(2) Feed the normalized 3D whole-body PET images and the 3D whole-body CT images jointly into the trained LR-DVF estimator [22] to directly predict the LR-DVF;

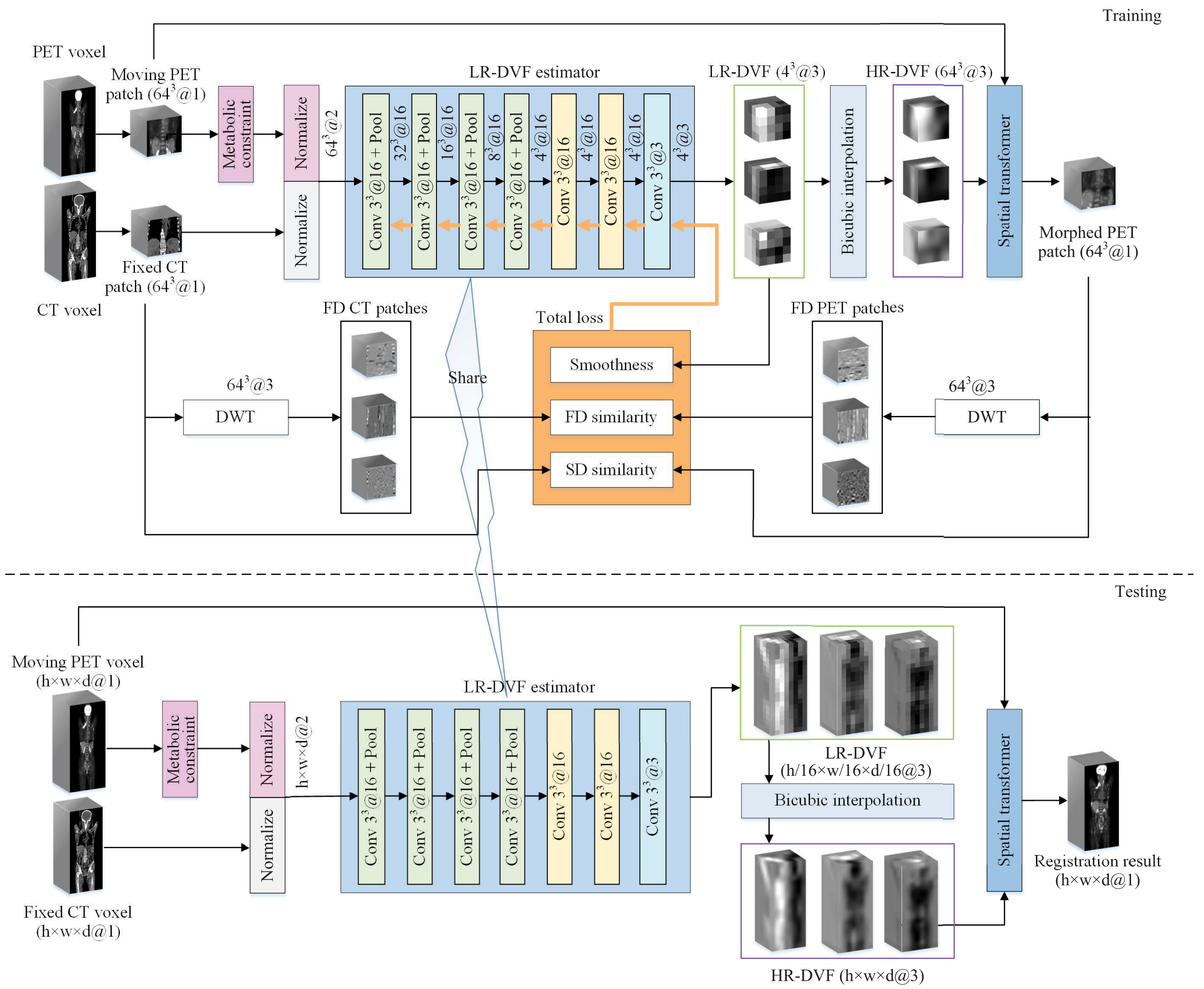


FIGURE 3. **The proposed 3D unsupervised PET-CT image registration network structure.**

(3) Perform bicubic interpolation on LR-DVF to generate HR-DVF;

(4) Feed the HR-DVF and original 3D whole-body PET images jointly into the spatial transformer for patient-level PET-CT image registration.

C. MULTI-DOMAIN SIMILARITY LOSS

The conventional voxel-based similarity measure is based on spatial domain and is suitable for a variety of registration tasks. However, the voxel intensity of hypermetabolic anatomical structures is usually low in CT images, making it difficult to correctly align the anatomical structures in PET-CT image registration tasks by using only spatial domain similarity measures.

To compensate for the misalignment by using voxel-based similarity measure in PET-CT image registration tasks, a frequency domain similarity measure is proposed in this paper. Though the voxel intensity of hypermetabolic regions in CT images is usually low, the contours of perfectly aligned anatomical structures in PET-CT images are still partially matched. Therefore, enhancing the sensitivity to contour information of hypermetabolic anatomical structures is expected to improve the registration accuracy. Specifically, the original CT image patches and the current registration results are transformed into frequency domain information by discrete wavelet transform (DWT) in the training process. Note that only high frequency information is preserved for the frequency domain similarity calculation, which means that the low frequency information is discarded in the process of loss calculation.

To improve the registration accuracy by using the context information of anatomical structures in PET-CT images, it is also needed to introduce a smoothness measure into the loss function. Specifically, the gradient of the LR-DVF tensor in horizontal and vertical directions is used to compose a differentiable voxel-level regular term [22], [25] in the training process, which is used to minimize the abnormal spatial transformation of adjacent voxels and ensure the smoothness of the registration.

Based on the above considerations, a multi-domain similarity loss function is proposed in this paper, which is developed for whole-body PET-CT image registration tasks. The coordinated multidomain (spatial domain and frequency domain) loss function includes the following parts. (1) Spatial domain similarity measure; (2) Frequency domain similarity measure; (3) Displacement vector field smoothness measure. Let M denotes a moving image, F denotes a fixed image, and the registration result is denoted by M_t, where t is the number of iterations in the training stage. Y_{SD} and Y_{FD} are the similarity measures in the spatial domain and the frequency domain, which are used to align the anatomical structures. The DVF smoothness measure $R(D)$ is used to minimize abnormal deformations. Therefore, the definition of loss function L is shown as follows.

$$L = -\lambda_1 \cdot Y_{SD}(F, M_t) - \lambda_2 \cdot Y_{FD}(F, M_t) + \lambda_3 \cdot R(D) \tag{2}$$

where the hyperparameters λ_i $(i = 1, 2, 3)$ are used to balance the similarity measures and the DVF smoothness measure $R(D)$ and D is the tensor representing LR-DVF. $Y_{SD}(F, M_t)$ and $Y_{FD}(F, M_t)$ are the similarity measures between fixed images and current registration results in the spatial domain and the frequency domain, respectively. Since the size of whole-body PET-CT images varies with the physical condition of patients, a similarity measure based on normalized cross correlation (NCC) is used in this paper. Compared with the similarity measures based on cross correlation, NCC does not depend on the size and contrast of an image [24]. The spatial domain similarity measure Y_{SD} and frequency domain similarity measure Y_{FD} can be calculated as

$$Y_{SD}(F, M_t) = NCC(F, M_t) \tag{3}$$

$$Y_{FD}(F, M_t) = NCC(WT(F), WT(M_t)) \tag{4}$$

where WT is the wavelet transform operation. As for the DVF smoothness, the L_1 norm of the LR-DVF partial derivative is used instead of the L_2 norm, which is inspired by the regular term in [25]. The regular term $R(D)$ can be calculated as

$$R(D) = \sum_{i \in \Omega} \|\nabla D_i\|_1 \tag{5}$$

where D_i is a voxel in the LR-DVF tensor, the gradient of D_i in the horizontal and vertical directions is calculated by using the gradient operation ∇, and Ω represents the range of the 3D input image (or image patch). Therefore, the total loss function of 3D MC-MDS Net can be calculated as

$$\begin{aligned} L = &-\lambda_1 \cdot NCC(F, M_t) \\ &- \lambda_2 \cdot NCC(WT(F), WT(M_t)) \\ &+ \lambda_3 \sum\nolimits_{i \in \Omega} \|\nabla D_i\|_1 \end{aligned} \tag{6}$$

where ∇D_i is the gradient of LR-DVF tensor.

III. EXPERIMENTS

A. DATASET

A dataset consisted of 170 whole-body PET-CT images is used for registration performance evaluation, of which 103 cases are male and the remaining 67 cases are female. ^{18}F-FDG is used as the radioactive tracer in the PET imaging process with a delay of 60 minutes. The spatial resolution of CT images and PET images in the horizontal direction is 0.976562mm and 5.46875mm, respectively. The experimental dataset is divided into two parts. (1) The training dataset which contains 3D PET-CT patches of size $64 \times 64 \times 64$; (2) The testing dataset which contains 3D whole-body PET-CT images of size $128 \times 128 \times n$ (n ranges from 215 to 351). The dataset generation process consists of the following steps.

(1) Load the 128×128 sized PET images from a selected patient and sort them by PET sequence numbers;

(2) Load the 512×512 sized CT images from the same patient and sort them by CT sequence numbers;

(3) Reconstruct the sorted 2D PET and CT images into a 3D PET image ($128 \times 128 \times n$) and a 3D CT image ($512 \times 512 \times n$);

(4) Convert the voxels in 3D CT images to Hu values and adjust the window level and the window width;

(5) Convert the voxels in 3D PET images to SUV;

(6) For the testing dataset generation, down-sample the 3D CT image from $512 \times 512 \times n$ to $128 \times 128 \times n$ and compose the 3D whole-body PET-CT image;

(7) For the training dataset generation, crop the 3D whole-body PET-CT image of each patient separately into patches of size $64 \times 64 \times 64$ with a stride of 32 voxels and the sampled patches are discarded if the sliding window exceeds the image boundary;

(8) Repeat step (1) to (7) for all patient cases in the dataset.

Finally, a total of 21,558 3D PET-CT image patches is generated from 170 3D whole-body PET-CT images and all generated images and patches are normalized before feeding into the network.

B. TRAINING PROCESSES AND HYPERPARAMETER SETTINGS

A 10-fold cross-validation is used to evaluate the performance of 3D MC-MDS Net and baseline methods (Section III-C). The dataset consisted of 170 PET-CT scans are divided into 10 parts by random seeds. In each part, image data of 17 patients (10%) are selected to compose the testing dataset and the remaining 153 PET-CT scans (90%) with their patches are used for the training process. The $64 \times 64 \times 64$ sized 3D PET-CT image patches are fed into the network for training, while $128 \times 128 \times n$ (n ranges from 215 to 351) sized 3D whole-body PET-CT images in the testing dataset are fed into the trained LR-DVF estimator for registration. To save memory usage for whole-body images, all tensors for the loss function calculation are released in the testing process. Since the registration results are spatially transformed from the original 3D PET images, which are not metabolic constrained, all voxels (SUV) larger than 5 are clipped to 5 in the visualized registration result and used for subsequent quantitative evaluation.

The hyperparameters of the metabolic constraint model (Eq. (1)), which are determined in advance, can be set by using the statistical data from Table 1 and Figure 2. First, most of the low metabolic regions (SUV < 1.8) in Figure 1 do not exhibit excessive distortion in the registration process. Therefore, set $\mu_1 = 1.8$, $k_1 = 1$, $b_1 = 0$ for keeping the original low metabolic information in PET images. Second, the registration distortions caused by native hypermetabolic anatomical structures such as brain and heart are visually different from the distortions caused by anatomical structures with pseudo-metabolic phenomenon (bladder). Thus, according the metabolic level distributions in Table 1, $\mu_2 = 20$ is set for separating different type of hypermetabolic anatomical structures. Then, $k_2 = 0.038$, $b_2 = 1.732$ are set according to the monotonically increasing property of the metabolic constraint model. Finally, $k_3 = 0.003$, $b_3 = 2.440$ are used to limit the output value of Eq. (1).

In [22], the loss function is constructed by the first term (space domain similarity) and the third term (regular term) in Eq. (6), i.e., $L = -\lambda_1 \cdot NCC(F, M_t) + \lambda_3 \cdot \sum_{i\in\Omega} \|\nabla D_i\|_1$. Since the goal of registration is to maximize the similarity between registered images and fixed images, the value of λ_1 is set to 1 for investigating the relationship between λ_3 and NCC. In the experiment, the value of λ_3 is chosen from range (0.01, 0.1) with a stride of 0.01 and the optimal value of λ_3 is 0.04 (Figure 4).

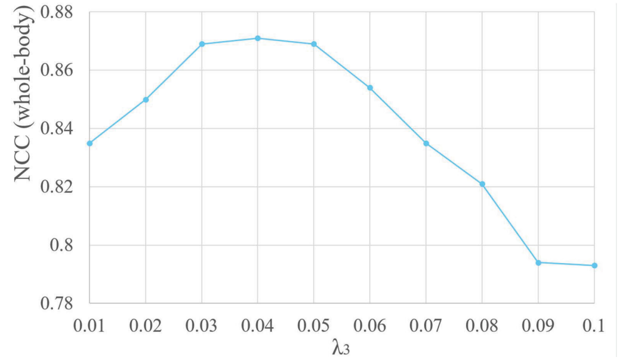


FIGURE 4. The relationship between regular term coefficient λ_3 and whole-body NCC.

In this paper, the relationship between spatial domain similarity coefficient (λ_1), frequency domain similarity coefficient (λ_2) and whole-body NCC is also discussed. In the experiment, the value of λ_3 is set to 0.04 according to the previous discussion and a grid searching is performed in range (0, 1.25) with a stride of 0.25. Based on the grid searching result in Figure 5, the parameter values corresponding to the highest whole-body NCC ($\lambda_1 = 1$, $\lambda_2 = 1$, $\lambda_3 = 0.04$) are set for optimal registration accuracy.

With regard to the convergence issue, an early stopping mechanism is used in the training process for network convergence, which monitors the changes of loss (Eq. (6)). As long as the network loss value continuously increases, the early stopping mechanism is triggered and the training process is completed.

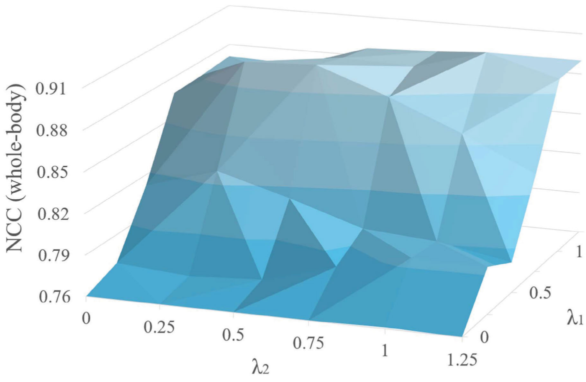


FIGURE 5. Grid searching result of spatial domain similarity coefficient (λ_1) and frequency domain similarity coefficient (λ_2) in Eq. (6).

Adam optimizer is used in the training process and the learning rate is 1.0×10^{-4} with 10,000 iterations. The proposed neural network is constructed using TensorFlow and the 3D bicubic interpolation operation is composed of two GPU-accelerated 2D bicubic interpolation operations. The DWT from spatial domain images to frequency domain information is implemented by the PyWavelet library and the 3D spatial transformer is implemented according to [23].

C. BASELINE METHODS AND ABLATION STUDIES

1) BASELINE METHODS

To evaluate and compare the registration accuracy of the proposed registration method for PET-CT images, four baseline methods are implemented: VoxelMorph [23], 2D deformable image registration network (2D DIR Net) [20], Advanced normalization tools (ANTs) [21] and 3D CNN with regular term (3D CNN-RT) [22].

(1) **VoxelMorph** [23]: VoxelMorph is a deformable image registration method that uses a U-Net based network structure to predict the deformation parameters for a pair of images. The registration is accomplished by feeding a moving image into a parameterized spatial transformer based on the predicted parameters. In this experiment, we applied an open source implementation of this method to whole-body PET-CT images registrations.[1] The hyperparameters ($\lambda = 1$) recommended by [23] are used after grid searching and the final registration is obtained through a single forward propagation.

(2) **2D DIR Net** [20]: 2D DIR Net is an unsupervised medical image registration method based on deep learning technology and its registration accuracy in MR images is better than traditional rigid registration methods with improved registration efficiency. Since 2D DIR Net is used for 2D image registration, it is impossible to input 3D images directly into this network. To ensure the fairness of the evaluation process, the 3D whole-body PET-CT images are sliced into 2D images

[1] https://github.com/voxelmorph/voxelmorph

of size 128 × 128 for training and testing and the warped 2D images of size 128 × 128 are stacked into 3D images for both visual and quantitative evaluation.

(3) **ANTs** [21]: The ANTs software is a conventional image registration toolkit. The similarity measure between a moving image and a fixed image is used to update deformation parameters in an iterative process to generate the registration result. The Symmetric normalization (SyN) implementation in the ANTs software toolkit is used for PET-CT image registration with default arguments, which applies both affine and elastic deformations to the moving image and uses the mutual information as a target function in the iterative process with adequate computational efficiency.

(4) **3D CNN-RT** [22]: 3D CNN-RT is a 3D unsupervised image registration method based on convolutional neural networks, which is consisted of a CNN regressor and a 3D spatial transformer [23]. Compared with conventional 2D registration methods, it makes full use of the spatial information contained in the 3D images and avoids over deformation in the registration process by introducing a smoothness measure of displacement vector field in the loss function.

Note that feature point based registration methods are suitable for structural images because the feature points (corner points, inflection points, etc.) of these images can be easily found. However, the PET images in PET-CT registration tasks are blurred and contain metabolic information rather than structural information. Therefore, not all of the feature points in PET images can be explicitly extracted and registration methods based on feature point are not included in baseline methods.

2) ABLATION STUDIES

To better demonstrate the effects of the metabolic constraint model and frequency domain similarity measure proposed in this paper on the registration accuracy of PET-CT images, three ablation studies (3D MC-MDS Net, 3D MC-MDS Net (w/o MDS), 3D MC-MDS Net (w/o MC)) are performed on the same dataset.

(1) **3D MC-MDS Net**: The proposed 3D MC-MDS Net includes a metabolic constraint model and introduces multi-domain similarity to construct loss function. This network can automatically learn the registration model of PET-CT images in an unsupervised manner. After training, the final registration results can be generated by a single feed-forward propagation.

(2) **3D MC-MDS Net (w/o MDS)**: 3D MC-MDS Net (w/o MDS) refers to the addition of the metabolic constraint model based on 3D CNN-RT (3D CNN-RT + MC), which is also a frequency domain similarity removed version of 3D MC-MDS Net. This experiment can verify the effects of frequency domain similarity for PET-CT image registration accuracy compared with conventional intensity-based similarity. Because the regular term of displacement vector field can improve the accuracy of PET-CT image registration task [22], the loss function in this experiment includes spatial domain similarity measure and regular term, which means that all the network structures in this study are the same as 3D MC-MDS Net except the loss function.

(3) **3D MC-MDS Net (w/o MC)**: 3D MC-MDS Net (w/o MC) refers to the deployment of multi-domain similarity based on 3D CNN-RT (3D CNN-RT + MDS), which is also a metabolic constraint model removed version of 3D MC-MDS Net. This experiment can verify the effect of the metabolic constraint model on abnormal deformations caused by hypermetabolic anatomical structures. In addition, the network output is truncated to a range of SUV (0, 5) for further visualization processes due to the high fluctuation level of SUV in the original PET images (see Table 1 and Figure 2), which is also used in [22]. Note that the network structure of this ablation study is the same as 3D MC-MDS Net except the metabolic constraint model.

D. EVALUATION CRITERIA

Due to the difficulty of obtaining voxel-level ground truth transformations for whole-body PET-CT images, the registration accuracy cannot be evaluated by directly calculating the mean distances of anatomical landmarks. Instead, normalized cross correlation [24], mutual information [26] and normalized mutual information [27] are used to evaluate the registration accuracy. The above registration accuracy evaluation measures are calculated as follows.

$$NCC(M,F) = \frac{\sum_{i=1}^{m}\sum_{j=1}^{n}(M_i - \bar{M})(F_j - \bar{F})}{\sigma(M)\sigma(F)} \tag{7}$$

$$MI(M,F) = \sum_{i=1}^{m}\sum_{j=1}^{n} P(M_i, F_j) log \frac{P(M_i, F_j)}{P(M_i)P^{\prime(}F_j)} \tag{8}$$

$$NMI(M,F) = \frac{2MI(M,F)}{-\sum_{i=1}^{m}P(M_i)logP(M_i) - \sum_{j=1}^{n}P'(F_j)logP'(F_j)} \tag{9}$$

where functions $\bar{M}$ and $\sigma(M)$ are the mean and standard deviation of all voxels in the image M, respectively. $P(M_i)$ denotes the probability of voxel M_i appearing in the image M. Similarly, $P'(F_j)$ denotes the probability of voxel F_j appearing in the image F. In addition, $P(M_i, F_j)$ is the joint probability of voxel pair (M_i, F_j).

To evaluate the registration accuracy of brain, the scores of brain region are also introduced by using the voxel data within the brain bounding box. Note that only 42 of the 170 whole-body PET-CT images in the dataset contains brain regions and are used for registration accuracy evaluation. The brain bounding box refers to the image slices between the uppermost part of the skull and the maxilla on the CT images.

In addition to comparing with the baseline methods, reference truth can be obtained by adjusting the pixel spacing of PET images [28]. Note that ground truth or ground truth transformations are not used in the training process. Specifically, the pixel spacing of PET image (5.46875mm) is adjusted to CT image (0.976562mm) for generating reference truth. Although this method cannot correct non-rigid deformation caused by breathing and only work for PET-CT images of the

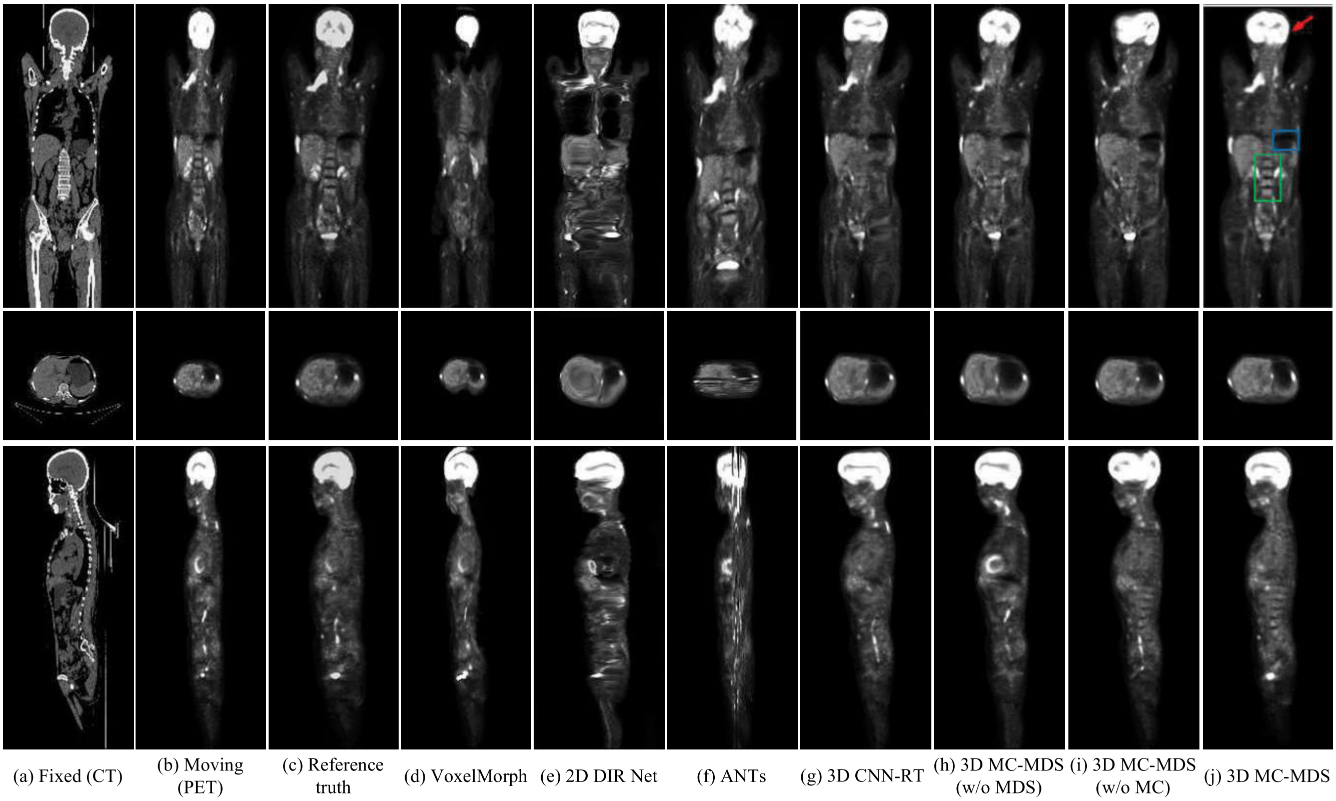


FIGURE 6. **Whole-body PET-CT image registration results. (a) Fixed image (CT). (b) Moving image (PET). (c) Reference truth. (d) VoxelMorph [23] registration result. (e) 2D DIR Net [20] registration result. (f) ANTs [21] registration result. (g) 3D CNN-RT [22] registration result. (h) 3D MC-MDS (w/o MDS) registration result. (i) 3D MC-MDS (w/o MC) registration result. (j) 3D MC-MDS registration result. Each registered 3D image is visualized in the coronal section (top), axial section (middle) and sagittal section (bottom).**

same patient, it is still suitable for generating reference truth. Note that all quantitative evaluation measures are calculated between the registration results and the corresponding reference truth.

IV. RESULTS AND DISCUSSIONS

The registration results of 3D PET-CT images are shown in Figure 6. The proposed method not only significantly improves the registration accuracy of high metabolic brain (red arrow), but also suppresses the mismatch of anatomical structures near the diaphragm (blue rectangular area) caused by breathing during imaging process. In addition, the registration accuracy in spine area (green rectangular area) is also improved.

In the brain region, VoxelMorph (Figure 6(d)), 2D DIR Net (Figure 6(e)), ANTs (Figure 6(f)), 3D CNN-RT (Figure 6(g)) and 3D MC-MDS (w/o MC) (Figure 6(i)) all generate excessive distortions, resulting in the loss of the contour and internal textures. Referring to the reference truth (Figure 6(c)), the 3D MC-MDS (w/o MDS) (Figure 6(h)) maintains the outline of the brain region but not internal textures, while the proposed method (Figure 6(j)) maintains most of the textures of the brain.

Near the diaphragm, VoxelMorph (Figure 6(d)), 2D DIR Net (Figure 6(e)) and ANTs (Figure 6(f)) give significantly distorted results, which reduces the registration accuracy. Since the reference truth (Figure 6(c)) does not compensate for the misalignment between the moving images (PET) and the fixed images (CT) caused by breathing in the region near the diaphragm, the registration accuracy is determined by using the anatomical structures contained in the fixed image (Figure 6(a)). 3D MC-MDS (w/o MDS) (Figure 6(h)) does not correct the non-rigid deformation mentioned above. Although the position of diaphragm is not completely warped to the fixed image (Figure 6(a)), 3D CNN-RT (Figure 6(g)), 3D MC-MDS (w/o MC) (Figure 6(i)) and 3D MC-MDS (Figure 6(j)) all correct the low metabolic region misalignment in PET images.

Near the spine, VoxelMorph (Figure 6(d)), 2D DIR Net (Figure 6(e)), 3D CNN-RT (Figure 6(g)) and 3D MC-MDS (w/o MDS) (Figure 6(h)) all lost most of the original textures in the moving PET images, while ANTs (Figure 6(f)), 3D MC-MDS (w/o MC) (Figure 6(i)) and 3D MC-MDS (Figure 6(j)) have higher registration accuracy in the region near the spine.

All registration results and their difference images with corresponding reference truth (coronal section) are shown in Figure 7. The lower the brightness of the difference images, the better the registration accuracy. Compared with the baseline methods, the difference image of 3D MC-MDS (Figure 7(r)) achieves lower errors. The difference image of 3D MC-MDS (w/o MDS) (Figure 7(p)) has the lowest

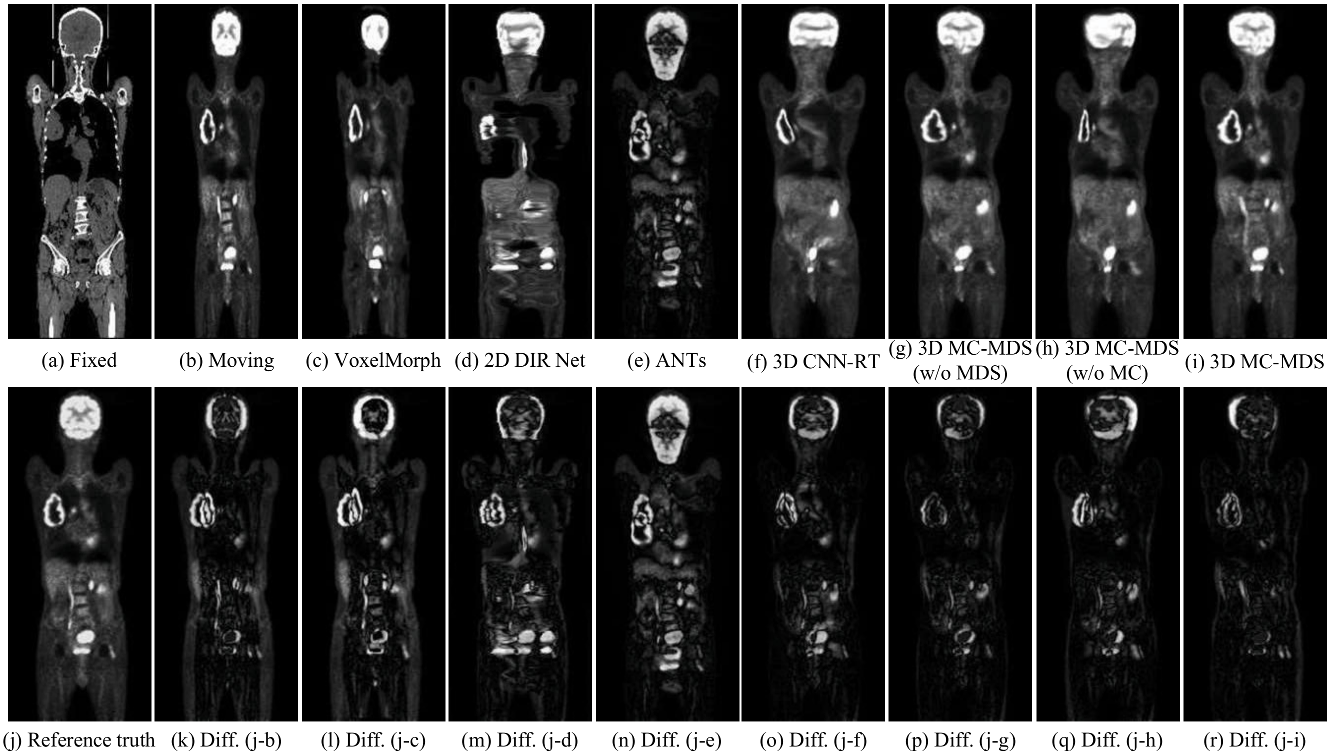


FIGURE 7. Difference images (k-r) between moving images (b-i) and the reference truth (j).

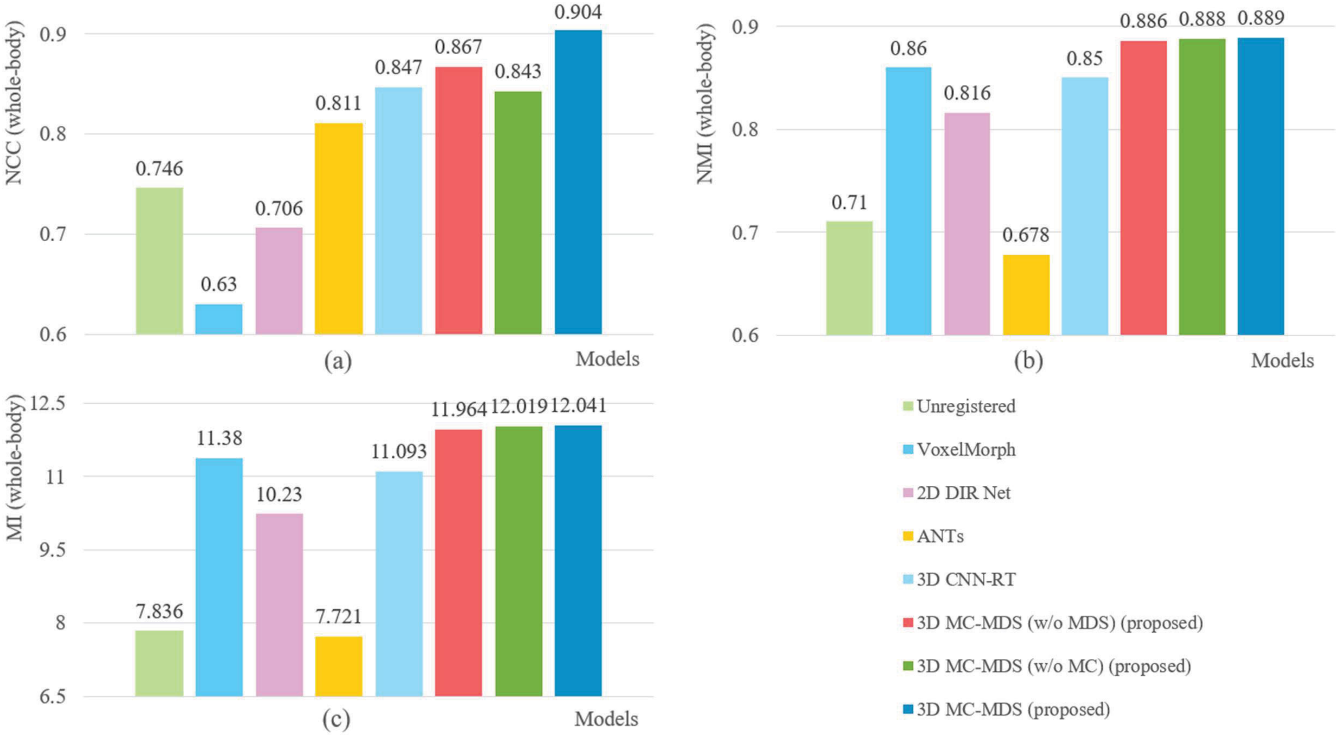


FIGURE 8. Quantitative evaluation results on 170 whole-body PET-CT images.

brightness in the brain but higher in other regions. The difference images of 3D CNN-RT (Figure 7(o)) and 3D MC-MDS (w/o MC) (Figure 7(q)) show more distortions in the high metabolic regions. Besides, the difference images of VoxelMorph (Figure 7(l)), 2D DIR Net (Figure 7(m)) and ANTs (Figure 7(n)) have more bright regions, indicating that the anatomical structures are excessively distorted by registration process.

The quantitative evaluations of the registration accuracy on 170 PET-CT images are shown in Figure 8 and Table 2.

TABLE 2. The quantitative evaluation of all registration methods.

Methods	NCC		MI		NMI		CPU/GPU sec.
	Whole-body	Brain	Whole-body	Brain	Whole-body	Brain	
Unregistered	0.746±0.035	0.749±0.021	7.836±0.551	6.500±0.458	0.710±0.027	0.690±0.019	-
VoxelMorph [23]	0.630±0.110	0.618±0.029	11.380±0.741	7.465±0.280	0.862±0.027	0.800±0.015	2.652±0.303
2D DIR Net [20]	0.706±0.069	0.833±0.035	10.230±0.454	10.206±0.579	0.816±0.018	0.864±0.020	2.677±0.027
ANTs [21]	0.811±0.042	0.783±0.060	7.721±0.548	6.248±0.399	0.678±0.027	0.636±0.018	62.370±8.110
3D CNN-RT [22]	0.847±0.042	0.823±0.033	11.093±0.489	10.490±0.620	0.850±0.019	0.877±0.020	**2.354±0.305**
3D MC-MDS (w/o MDS) (proposed)	0.867±0.038	0.867±0.034	11.964±0.467	10.758±0.651	0.886±0.015	0.889±0.021	2.573±0.349
3D MC-MDS (w/o MC) (proposed)	0.843±0.057	0.813±0.040	**12.109±0.467**	10.792±0.664	0.888±0.015	0.889±0.022	2.575±0.359
3D MC-MDS (proposed)	**0.904±0.028**	**0.895±0.027**	12.041±0.467	**10.809±0.652**	**0.889±0.015**	**0.891±0.021**	2.569±0.352

Our methods obtain the best results in NCC, MI and NMI scores, indicating that 3D MC-MDS outperformed the other methods in terms of registration accuracy. The execution time per patient for registration is also listed in Table 2. Compared with the iteration-based methods, the proposed method accelerates registration process by applying single feed-forward propagation scheme.

In order to verify the bidirectional reversibility of the proposed method, a CT-to-PET experiment is performed and the quantitative evaluation results are shown in Table 3.

TABLE 3. The quantitative evaluation of the proposed method in both direction (Avg.).

Evaluation index	PET-to-CT	CT-to-PET
NCC	**0.902**	0.723
MI	**11.973**	2.194
NMI	**0.886**	0.684

The registration performance of the PET-to-CT mode in Table 3 is better than that of the CT-to-PET mode. We believe that the blurred PET images are regarded as fixed images in the CT-to-PET mode, which leads to relatively large errors in DVF and reduces the registration accuracy compared to the PET-to-CT mode. Thus, the PET-to-CT mode is used in this work.

The proposed method and DenseRegNet [30] both perform the 3D PET-CT image registration, but these two methods are different at (1) The loss function of DenseRegNet includes a 2D DIR Net [20] derived similarity measure and a regular term, which is defined in the spatial domain. Different from DenseRegNet, the proposed loss function consists of a spatial domain similarity measure, a frequency domain similarity measure and a regular term; (2) A novel metabolic constraint function is defined and applied to PET images in the proposed method to reduce the excessive distortion in hypermetabolic anatomical structures. The quantitative evaluation of the proposed method and DenseRegNet is shown in Table 4 and Table 5.

TABLE 4. The quantitative evaluation of the proposed method and DenseRegNet in whole-body regions (Avg.).

Evaluation index	DenseRegNet [30]	Proposed
NCC	**0.909**	0.904
MI	**12.086**	12.041
NMI	**0.890**	0.889

TABLE 5. The quantitative evaluation of the proposed method and DenseRegNet in hypermetabolic brain regions (Avg.).

Evaluation index	DenseRegNet [30]	Proposed
NCC (brain)	0.824	**0.895**
MI (brain)	10.415	**10.809**
NMI (brain)	0.782	**0.891**

In Table 4, the whole-body registration accuracy of the proposed method is slightly lower than DenseRegNet. However, the proposed method obtains higher scores in the hypermetabolic brain region (Table 5).

The above work shows that the non-rigid deformations caused by breathing in PET-CT imaging process lead to systematic differences of anatomical structure alignments in corresponding PET-CT images, which cannot be compensated by rigid registration. Although there is no ground truth transformation for network training, the proposed method (3D MC-MDS Net) successfully learns the PET-CT image registration model by an unsupervised learning scheme. A metabolic constraint model is used to suppress the registration error of hypermetabolic region and a multi-domain similarity function is used to maintain the texture of the anatomical structure in the non-rigid registration process, which significantly improves the registration accuracy on whole-body PET-CT images.

V. CONCLUSIONS

In this paper, we propose an unsupervised registration network (3D MC-MDS Net) for 3D whole-body PET-CT images, which avoids the usage of ground truth transformations for network training. In the proposed method, a metabolic constraint model is applied to suppress the registration distortion caused by hypermetabolic regions in PET images. In addition, we also introduce a multi-domain similarity based loss function to improve the registration accuracy. In the testing process, a dataset consisted of 170 whole-body PET-CT images are used for registration accuracy evaluation. The visual and quantitative evaluation results demonstrate that the proposed method with metabolic constraint model and multi-domain similarity measures significantly improves the registration accuracy on 3D whole-body PET-CT images compared with baseline methods, which is expected to improve the sensitivity and specificity of early cancer diagnosis systems.

REFERENCES

[1] R. Bar-Shalom, N. Yefremov, L. Guralnik, D. Gaitini, A. Frenkel, A. Kuten, H. Altman, Z. Keidar, and O. Israel, "Clinical performance of PET/CT in evaluation of cancer: Additional value for diagnostic imaging and patient management," *J. Nucl. Med.*, vol. 44, no. 8, pp. 1200–1209, 2003.

[2] D. Mattes, D. R. Haynor, H. Vesselle, T. K. Lewellen, and W. Eubank, "PET-CT image registration in the chest using free-form deformations," *IEEE Trans. Med. Imag.*, vol. 22, no. 1, pp. 120–128, Jan. 2003.

[3] G. W. Goerres, E. Kamel, T.-N.-H. Heidelberg, M. R. Schwitter, C. Burger, and G. K. von Schulthess, "PET-CT image co-registration in the thorax: Influence of respiration," *Eur. J. Nucl. Med. Mol. Imag.*, vol. 29, no. 3, pp. 351–360, Mar. 2002.

[4] R. Shekhar, V. Walimbe, S. Raja, V. Zagrodsky, M. Kanvinde, G. Wu, and B. Bybel, "Automated 3-dimensional elastic registration of whole-body pet and ct from separate or combined scanners," *J. Nucl. Med.*, vol. 46, no. 9, pp. 1488–1496, 2005.

[5] S. Sultana, D. Y. Song, and J. Lee, "Deformable registration of PET/CT and ultrasound for disease-targeted focal prostate brachytherapy," *J. Med. Imag.*, vol. 6, no. 3, Sep. 2019, Art. no. 035003.

[6] S. Sultana, D. Y. Song, and J. Lee, "A deformable multimodal image registration using PET/CT and TRUS for intraoperative focal prostate brachytherapy," in *Proc. Med. Imag., Image-Guided Procedures, Robotic Interventions, Modeling*, Mar. 2019, Art. no. 109511.

[7] A. Collignon, F. Maes, D. Delaere, D. Vandermeulen, P. Suetens, and G. Marchal, "Automated multi-modality image registration based on information theory," *Inf. Process. Med. Imag.*, vol. 3, no. 6, pp. 263–274, 1995.

[8] W. M. Wells, III, P. Viola, H. Atsumi, S. Nakajima, and R. Kikinis, "Multimodal volume registration by maximization of mutual information," *Med. Image Anal.*, vol. 1, no. 1, pp. 35–51, 1996.

[9] F. Maes, A. Collignon, D. Vandermeulen, G. Marchal, and P. Suetens, "Multimodality image registration by maximization of mutual information," *IEEE Trans. Med. Imag.*, vol. 16, no. 2, pp. 187–198, Apr. 1997.

[10] M. A. Viergever, J. B. A. Maintz, S. Klein, K. Murphy, M. Staring, and J. P. W. Pluim, "A survey of medical image registration–under review," *Med. Image Anal.*, vol. 33, pp. 140–144, Oct. 2016.

[11] J. P. W. Pluim, J. B. A. Maintz, and M. A. Viergever, "Mutual-information-based registration of medical images: A survey," *IEEE Trans. Med. Imag.*, vol. 22, no. 8, pp. 986–1004, Aug. 2003.

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

[13] A. Sedghi, J. Luo, A. Mehrtash, S. Pieper, C. M. Tempany, T. Kapur, P. Mousavi, and W. M. Wells, III, "Semi-supervised deep metrics for image registration," 2018, *arXiv:1804.01565*. [Online]. Available: http://arxiv.org/abs/1804.01565

[14] R. Liao, S. Miao, P. de Tournemire, S. Grbic, A. Kamen, T. Mansi, and D. Comaniciu, "An artificial agent for robust image registration," in *Proc. 31st AAAI Conf. Artif. Intell.*, 2017, pp. 4168–4175.

[15] S. Miao, Z. J. Wang, Y. Zheng, and R. Liao, "Real-time 2D/3D registration via CNN regression," in *Proc. IEEE 13th Int. Symp. Biomed. Imag. (ISBI)*, Apr. 2016, pp. 1430–1434.

[16] G. Wu, M. Kim, Q. Wang, B. C. Munsell, and D. Shen, "Scalable high-performance image registration framework by unsupervised deep feature representations learning," *IEEE Trans. Biomed. Eng.*, vol. 63, no. 7, pp. 1505–1516, Jul. 2016.

[17] M. Jaderberg, K. Simonyan, and A. Zisserman, "Spatial transformer networks," in *Proc. Adv. Neural Inf. Process. Syst.*, 2015, pp. 2017–2025.

[18] G. Haskins, U. Kruger, and P. Yan, "Deep learning in medical image registration: A survey," 2019, *arXiv:1903.02026*. [Online]. Available: http://arxiv.org/abs/1903.02026

[19] J. Lv, M. Yang, J. Zhang, and X. Wang, "Respiratory motion correction for free-breathing 3D abdominal MRI using CNN-based image registration: A feasibility study," *Brit. J. Radiol.*, vol. 91, Jan. 2018, Art. no. 20170788.

[20] B. D. de Vos, F. F. Berendsen, M. A. Viergever, M. Staring, and I. Išgum, "End-to-end unsupervised deformable image registration with a convolutional neural network," in *Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support*. Québec City, QC, Canada: Springer, 2017, pp. 204–212.

[21] B. B. Avants, N. J. Tustison, G. Song, P. A. Cook, A. Klein, and J. C. Gee, "A reproducible evaluation of ANTs similarity metric performance in brain image registration," *NeuroImage*, vol. 54, no. 3, pp. 2033–2044, Feb. 2011.

[22] H. Jiang, H. Yu, X. Zhou, H. Kang, Z. Wang, T. Hara, and H. Fujita, "Learning 3D non-rigid deformation based on an unsupervised deep learning for PET/CT image registration," in *Proc. Med. Imag., Biomed. Appl. Mol., Struct., Funct. Imag.*, vol. 10953, Mar. 2019, Art. no. 109531.

[23] A. V. Dalca, G. Balakrishnan, J. Guttag, and M. R. Sabuncu, "Unsupervised learning for fast probabilistic diffeomorphic registration," in *Proc. Int. Conf. Med. Image Comput. Comput.-Assist. Intervent.* Granada, Spain: Springer, 2018, pp. 729–738.

[24] J.-C. Yoo and T. H. Han, "Fast normalized cross-correlation," *Circuits, Syst. Signal Process.*, vol. 28, no. 6, p. 819, 2009.

[25] H. Li and Y. Fan, "Non-rigid image registration using fully convolutional networks with deep self-supervision," 2017, *arXiv:1709.00799*. [Online]. Available: http://arxiv.org/abs/1709.00799

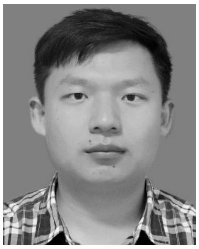
[26] P. Viola and W. M. Wells, III, "Alignment by maximization of mutual information," *Int. J. Comput. Vis.*, vol. 24, no. 2, pp. 137–154, 1997.

[27] P. A. Estevez, M. Tesmer, C. A. Perez, and J. M. Zurada, "Normalized mutual information feature selection," *IEEE Trans. Neural Netw.*, vol. 20, no. 2, pp. 189–201, Feb. 2009.

[28] Y. E. Erdi, S. A. Nehmeh, T. Pan, A. Pevsner, K. E. Rosenzweig, G. Mageras, E. D. Yorke, H. Schoder, W. Hsiao, and O. D. Squire, "The CT motion quantitation of lung lesions and its impact on PET-measured SUVs," *J. Nucl. Med.*, vol. 45, no. 8, pp. 1287–1292, 2004.

[29] O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," in *Proc. Int. Conf. Med. Image Comput. Comput.-Assist. Intervent.* Munich, Germany: Springer, 2015, pp. 234–241.

[30] H. Kang, H. Jiang, X. Zhou, H. Yu, T. Hara, H. Fujita, and Y.-D. Yao, "An optimized registration method based on distribution similarity and DVF smoothness for 3D PET and CT images," *IEEE Access*, vol. 8, pp. 1135–1145, 2020.



HENGJIAN YU is currently pursuing the M.S. degree with the Software College, Northeastern University, China. His research interests include medical image analysis and machine learning.

HUIYAN JIANG received the B.S. degree from the Department of Mathematics, Bohai University, China, in 1986, and the M.Sci. degree in computer application and the Ph.D. degree in control theory and control engineering from Northeastern University, China, in 2000 and 2009, respectively. From October 2001 to September 2002, she was a Visiting Scholar with Gifu University, Japan, where she carried out the research in image processing and medical image computer-aided diagnosis (CAD) technology. She is currently a Professor and the Director of the Department of Digital Media Technology, Software College, Northeastern University, China. Her main research interests are focus on the digital image processing and analysis, pattern recognition, 3D visualization, 3D video processing, artificial intelligence, and medical image computer-aided diagnosis (CAD). She is a Council Member of the 3D Images Technology Association, China Society of Image and Graphics.

XIANGRONG ZHOU received the M.S. and Ph.D. degrees in information engineering from Nagoya University, Japan, in 1997 and 2000, respectively. From 2000 to 2002, he continued his research in medical image processing as a Postdoctoral Researcher at Gifu University, Japan, where he is currently an Assistant Professor at the Department of Electrical, Electronic and Computer Engineering. His research interests include medical image analysis, medical image visualization, and pattern recognition.

TAKESHI HARA received the B.S. and M.S. degrees from the Faculty of Engineering, Gifu University, Japan, in 1994 and 1995, respectively, and the Ph.D. degree from Gifu University, in 2000. In 1995, he was a Research Assistant at the Faculty of Engineering, Gifu University, where he became an Associate Professor, in 2001 and 2017. He was also an Associate Professor at the Department of Intelligent Image Information, Gifu University. He was a Visiting Associate Professor at the Department of Radiology, The University of Chicago, from 2008 to 2009. He has been a Professor at the Faculty of Engineering and a Chair of the Artificial Intelligence Research Promotion Center, Medical Research Department, Gifu University, since 2019.

YU-DONG YAO (Fellow, IEEE) received the B.Eng. and M.Eng. degrees in electrical engineering from the Nanjing University of Posts and Telecommunications, Nanjing, China, in 1982 and 1985, respectively, and the Ph.D. degree in electrical engineering from Southeast University, Nanjing, in 1988. From 1987 to 1988, he was a Visiting Student at Carleton University, Ottawa, Canada. From 1989 to 2000, he was with Carleton University, Spar Aerospace Ltd., Montreal, Canada, and Qualcomm Inc., San Diego, USA. Since 2000, he has been with the Stevens Institute of Technology, Hoboken, USA, where he is currently a Professor and a Chair of the Department of Electrical and Computer Engineering. He holds one Chinese patent and 13 U.S. patents. His research interests include wireless communications, machine learning and deep learning techniques, and healthcare and medical applications. For his contributions to wireless communications systems, he was elected a Fellow of the National Academy of Inventors, in 2015, and the Canadian Academy of Engineering, in 2017. He served as an Associate Editor for the IEEE Communications Letters, from 2000 to 2008, and the IEEE Transactions on Vehicular Technology, from 2001 to 2006, and as an Editor for the IEEE Transactions on Wireless Communications, from 2001 to 2005.

HIROSHI FUJITA received the B.S. and M.S. degrees in electrical engineering from Gifu University, Japan, in 1976 and 1978, respectively, and the Ph.D. degree from Nagoya University, in 1983. In 1978, he became a Research Associate at the Gifu National College of Technology, where he also became an Associate Professor, in 1986. He was a Visiting Researcher at the K. Rossmann Radiologic Image Laboratory, The University of Chicago, from 1983 to 1986. He became an Associate Professor at the Faculty of Engineering, Gifu University, in 1991, where he also became a Professor, in 1995. He has been a Professor and a Chair of intelligent image information at the Graduate School of Medicine, Gifu University, since 2002. He is currently a Research Professor at Gifu University, and also a Visiting Professor at Zhengzhou University, China.