IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received November 8, 2019, accepted December 2, 2019, date of publication December 20, 2019, date of current version January 3, 2020.

Digital Object Identifier 10.1109/ACCESS.2019.2961268

An Optimized Registration Method Based on Distribution Similarity and DVF Smoothness for 3D PET and CT Images

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

[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-shi 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 in part by the National Natural Science Foundation of China under Grant 61872075, and in part by the Grant-in-Aid for Scientific Research on Innovative Areas (Multidisciplinary Computational Anatomy), MEXT, Japan, under Grant 26108005.

ABSTRACT A fusion image combining both anatomical and functional information obtained by registering medical images of two different modalities, Positron Emission Tomography (PET) and Computed Tomography (CT), is of great significance for medical image analysis and diagnosis. Medical image registration relies on similarity measure which is low between PET/CT image voxels and therefore PET/CT registration is a challenging task. To address this issue, this paper presents an unsupervised end-to-end method, DenseRegNet, for deformable 3D PET/CT image registration. The method consists of two stages: (1) predicting 3D displacement vector field (DVF); and (2) registering 3D image. In the 3D DVF prediction stage, a two-level similarity measure together with a deformation regularization is proposed as loss function to optimize network training.In the image registration stage, a resampler and a spatial transformer are utilized to obtain the registration results. In this paper, 663 pairs of Uptake Value (SUV) and Hounsfield Unit (Hu) patches of 106 patients, 227 pairs of SUV and Hu patches of 35 patients and 259 pairs of SUV and Hu patches of 35 patients are randomly selected as training, validation and test set, respectively. Normalized cross correlation (NCC), intersection over union (IoU) of liver bounding box and euclidean distance (ED) on landmark points are used to evaluate the registration results. Experiment results show that the proposed method, DenseRegNet, achieves the best results in terms of liver bounding box IoU and ED, and the second highest value of NCC. For a trained model, given a new pair of PET/CT images, the registration result can be obtained with only one forward calculation within 10 seconds. Through qualitative and quantitative analyses, we demonstrate that, compared with other deep learning registration models, the proposed DenseRegNet achieves improved results in the challenging deformable PET/CT registration task.

INDEX TERMS PET/CT registration, unsupervised learning, two-level similarity measure, deformation regularization.

I. INTRODUCTION

Multi-modality medical image registration technology is currently a research hotspot in medical image processing, and is of great significance for clinical diagnosis and treatment. Modern medical imaging technologies emerge in an endless stream [1], including structural images such as CT, Magnetic Resonance Imaging (MRI) and functional images such as PET, Functional Magnetic Resonance Imaging (FMRI), Single-Photon Emission Computed Tomography (SPECT). Structural images are used to analyze the anatomy of the lesion, while functional images are often used to diagnose the location of subtle lesions. In clinical diagnosis, PET/CT images are usually fused to make pathological structure and

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

functional metabolism of the lesion simultaneously on one image, which can better diagnose the disease and understand the pathological details [2]. However, spatial differences are caused due to different imaging mechanisms, including not only overall patient positioning and movement but also the involuntary and uncontrollable motion of internal organs. Registration of whole-body PET and CT images is a prerequisite for their meaningful fusion [3], [4].

Iterative optimization based traditional registration methods have been extensively studied in 3D medical image registration. The most widely used method is to translate the registration problem into an optimization problem to minimize the cost function [5]. Commonly used cost functions include mean square error (MSE) [6], mutual information (MI) [7], normalized mutual information (NMI) [8], normalized cross-correlation (NCC) [9] and gradient correlation (GC) [10]. These metrics compare images directly at the pixel level without understanding the higher-level structure in the image. Although there are global optimization methods such as simulated annealing algorithms [11] and genetic algorithms [12], they require comprehensive sampling of the parameter space, which leads to excessive computational costs. Few attempts have been made to seek heuristic semi-global optimization for a proper balance between robustness and computational cost [13].

Some work has been done to use a learning-based approach to perform other medical image registration tasks. References [14] and [15] used deep learning to learn end-to-end supervised rigid registration. Reference [16] proposed an end-to-end supervised RegNet network architecture to address non-rigid 3D image registration, which directly predict 3D non-rigid DVF for given fixed and moving images. However, in these supervised registration tasks, reference standards are often expensive and difficult to obtain.

Therefore, we hope to propose a novel deep learning method for ill-posed deformable PET/CT registration tasks to overcome these shortcomings while maintaining competitive registration performance.

Recently, [17] proposed a spatial transformer network (STN), which can be inserted into existing convolutional architectures, giving neural networks the ability to actively spatially transform feature maps without any extra training supervision or modification to the optimization process. By training the entire network end-to-end, the embedded STN predicts the best alignment to solve some particular tasks. Inspired by spatial transformer network, [18] successfully applied it to the field of medical imaging for the first time, enabling unsupervised learning based registration during the training phase. The spatial transformation layer is directly connected to the CNN, and the obtained deformation field is used to warp the moving image to obtain a registration result. Referring to this registration mode, deep CNNs in conjunction with STNs have been proposed recently to learn prediction models for image registration from pairs of fixed and moving images in an unsupervised fashion [19]–[21].

Up to date, some work has been proposed to solve the problem of PET/CT registration [21]–[25]. References [22] and [23] proposed a three-dimensional deformable PET/CT registration algorithm based on mutual information similarity measure. Reference [24] proposed a multi-thread registration method based on contour point cloud for 3D whole-body PET and CT images. Reference [25] proposed a B-spline Free Form Deformation based algorithm to describe both images and deformation field, for registration of thoracic PET and CT data. All these traditional PET/CT image registration methods perform registration tasks by iteratively optimizing the similarity measure, resulting in high computational cost. For each pair of images to be registered that have not been seen, these registration methods iteratively optimize the cost function from the beginning, which severely limits the registration speed and does not exploit the correlation between the image data distributions from the same data set at all. Reference [21] proposed a deep learning method for automatic image registration between 3D PET and CT images. But there still exists two problems to be improved: over-deforming of PET images and unsatisfactory results for organ-level registration.

We propose an unsupervised end-to-end image registration method for effective and efficient 3D PET and CT image registration. The proposed method directly predicts the 3D displacement vector field for a given fixed and moving image, and then obtains the registered image by spatial transformation. By sharing the same parameters for the data set, the process learns a generic representation that can register any new pairs from the same distribution. The training process of the network is unsupervised and does not require pre-labeled displacement vector field reference standards. We propose a two-level similarity measure together with a deformation regularization for the PET/CT registration task. Since there is no well-defined evaluation method for non-rigid PET/CT registration tasks according to previous studies in literature, we use NCC, liver bounding box IoU and euclidean distance on landmark points to evaluate the registration results. Compared with other registration models [18]–[21], [24], we demonstrate that the proposed method achieves improved performance results in the challenging deformable PET/CT registration task. We reduce a patient's PET/CT registration time from the previous hour period to less than 10 seconds.

Different from the prior work [21], several improvements has been made in three aspects of network structure, loss function and evaluation metrics in this paper. We use a ''DenseNet''-based architecture as the DVF regressor. Maximum Mean Discrepancy (MMD) is introduced into the loss function to measure the similarity in another level by mapping PET and CT images into a higher dimensional space. We additionally use liver bounding box IoU and euclidean distance on landmark points to evaluate the anatomical registration results.

The contributions of this paper are threefold.

(1) An unsupervised end-to-end medical image registration framework is built for 3D deformable PET/CT registration.

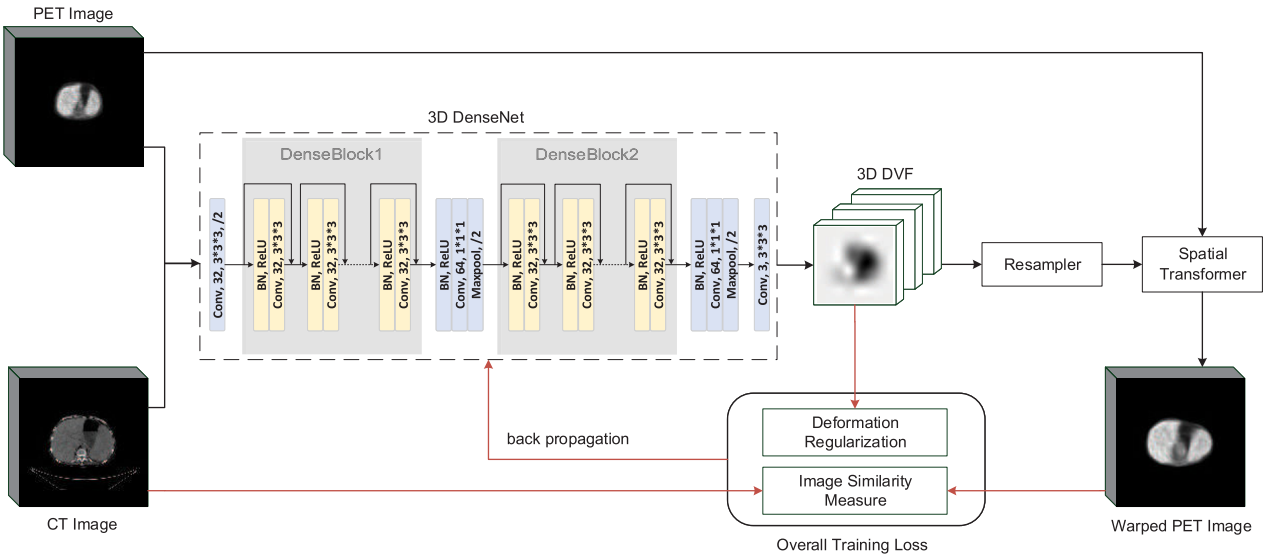


FIGURE 1. The architecture of our DenseRegNet. An illustration of the training strategy of the proposed unsupervised end-to-end registration method, where the red lines indicates data flows only required in training. The 3D DenseNet consists of two DenseBlocks and all operations are implemented in a 3D manner.

(2) A two-level similarity measure together with a deformation regularization is proposed for the PET/CT registration task.

(3) In addition to NCC, two anatomical metrics are utilized to evaluate registration results.

II. METHODS

For PET/CT registration, given a pair of fixed CT image I_F and moving PET image I_M, the task is to seek a spatial transformation that establishes pixel/voxel-wise spatial correspondence between the two images. Since the spatial correspondence can be gauged with a surrogate measure, such as an image intensity similarity measure between the fixed and the transformed moving images, the image registration task can be formulated as an optimization problem to optimize a spatial transformation that maximizes the image similarity measure between the fixed image and the transformed moving image. For non-rigid image registration, the spatial transformation is often characterized by a deformation field D_V that encodes displacement vector field (DVF) between spatial coordinates of I_F and their counterparts in I_M. As the PET/CT image registration problem is an ill-posed problem, regularization techniques are usually used to obtain smooth and anatomically plausible spatial transformations. The optimization based image registration problem is typically formulated as:

$$\hat{\phi} = \arg\min_{D_V} \mathcal{L}(I_F, I_M, D_V) \tag{1}$$

A. AN UNSUPERVISED 3D PET/CT REGISTRATION METHOD

We propose an unsupervised end-to-end image registration method, DenseRegNet, for effective and efficient 3D PET/CT registration. Figure 1 shows an overview of the architecture of our proposed DenseRegNet. DenseRegNet mainly consists of two components, a DenseNet for predicting 3D displacement vector fields and a Spatial Transformer for warping 3D images. DenseRegNet receives pre-processed 3D moving PET and fixed CT image patches I_M and I_F with the size of $128\times128\times64$ as input, and a displacement vector field D_V that consists of three $16\times16\times8$ sized patches is obtained through DenseNet regression to represent the displacement of the moving image patch I_M in the x, y, and z directions respectively. Then a Spatial Transformer is utilized to interpolate the 3D DVF to the same resolution with the input patch, and the registration result is obtained by warping the moving patch. DenseRegNet utilizes a similarity metric to learn and optimize the registration tasks end-to-end in an unsupervised manner.

We use a 3D DenseNet to predict 3D DVF. The network receives I_M and I_F as its input. A $3\times3\times3$ conv layer is utilized to extract shallow features and reduce dimensions, followed by two DenseBlocks, each of which contains C convolutional layers, and the number of convolution kernels is G (i.e. growth rate). Each DenseBlock is followed by BN, ReLU, a $1\times1\times1$ convolution layer and a max-pooling layer. Finally, a $3\times3\times3$ convolutional layer is used to regress a displacement vector field with the size of $16\times16\times8\times3$.

We summarize several advantages of the proposed method DenseRegNet, illustrated in Figure 1. First, DenseRegNet is an unsupervised end-to-end framework and does not require any pre-labeled reference standards for training process. Second, the 3D DenseNet preserves the maximum information flow between layers by a densely-connected mechanism and hence eases the network training. Meanwhile, it avoids learning redundant feature maps by encouraging feature

reuse. Third, we use a 16×16×8×3 displacement vector field to control the deformation of a 128×128×64 sized image (grid spacing of 8 voxels) and hence the network requires fewer parameters to achieve high performance.

B. TWO-LEVEL SIMILARITY MEASURE TOGETHER WITH DEFORMATION REGULARIZATION FOR PET/CT REGISTRATION

Direct use of traditional similarity measure between targets, such as those based on NCC and MI, are not appropriate for measuring image similarity in the context of PET/CT registration. On one hand, they do not consider the similarity of anatomical structure or data distribution in higher dimensions. They only focus on global information and ignore fine-grained local information (e.g. anatomical organs) which is crucial in multi-modal registration tasks. On the other hand, since some tissues or organs are highlighted in PET images while having lower contrast in CT images, using a single similarity measure without constraining the registration will over-deform the image and lead to a poor registration result.

To address the above problems, we propose a two-level similarity measure together with a deformation regularization for the PET/CT registration task. First, NCC is used to measure the similarity of voxels at the global level. Second, MMD is used to measure the similarity of data distributions at the higher dimensional level. And finally, DVF smoothness is used as a deformation regularization to limit the degree of deformation.

The total loss function $\mathcal{L}_{total}$ in this model consists of three parts mentioned above, as shown in Eq. (2):

$$\mathcal{L}_{total} = -\mathcal{L}_{NCC}(I_F, I_M{}') - \lambda_1 \mathcal{L}_{DS}(I_F, I_M{}') + \lambda_2 \mathcal{L}_{DR}(D_V) \tag{2}$$

where $\mathcal{L}_{NCC}(I_F, I_M{}')$ and $\mathcal{L}_{DS}(I_F, I_M{}')$ denotes the normalized cross correlation and the distribution similarity between the fixed and moving image patches after registration, respectively; $\mathcal{L}_{DR}(D_V)$ denotes the deformation regularization. I_F denotes the fixed CT image patch, and $I_M{}'$ denotes the moving PET image patch after registration, which can be obtained by Eq. (3):

$$I_M{}' = I_M \circ D_V \tag{3}$$

where I_M denotes the moving PET image patch before registration, $\circ$ denotes the transformation of the corresponding elements of I_M based on D_V. In this paper, DenseRegNet optimizes the total loss function $\mathcal{L}_{total}$ during the training process.

NCC is used to describe the degree of correlation between two targets, that is, the similarity between the targets [9]. We use NCC to measure the global similarity between moving and fixed image patches, as shown in Eq. (4):

$$\mathcal{L}_{NCC}(X, Y) = \frac{\sum(X - \bar{X}) \cdot \sum(Y - \bar{Y})}{\sigma(X) \cdot \sigma(Y)} \tag{4}$$

where $\bar{X}$ and $\bar{Y}$ denote the mean of the image X and Y, while $\sigma(X)$ and $\sigma(Y)$ denote the standard deviation of the image X and Y.

1) DISTRIBUTION SIMILARITY BASED REGISTRATION OPTIMIZATION

In domain adaptation, Maximum Mean Discrepancy has been widely adopted as a discrepancy metric between the distributions of source and target domains [26], [27]. It is mainly used to measure the distance of two different but related data distributions. In this paper, we calculate MMD distance as the distribution similarity (DS) measure, which is defined as Eq (5):

$$\mathcal{L}_{DS}(X, Y) = \| \frac{1}{n} \sum_{i=1}^{n} \phi(x_i) - \frac{1}{m} \sum_{j=1}^{m} \phi(y_j) \|_H^2 \tag{5}$$

where X and Y are the two distributions to be calculated, and H indicates that the distance is measured by $\phi(\cdot)$ mapping the data into the Reproducing Kernel Hilbert Space (RKHS). Expand Eq. (5) to obtain Eq. (6):

$$\mathcal{L}_{DS}(X, Y) = \| \frac{1}{n^2} \sum_{i,i'}^{n} \phi(x_i)\phi(x_i{}') - \frac{2}{mn} \sum_{i,j}^{n} \phi(x_i)\phi(y_j) + \frac{1}{m^2} \sum_{j,j'}^{n} \phi(y_j)\phi(y_j{}') \|_H^2 \tag{6}$$

By contacting the kernel function $k(*)$ in the support vector machine (SVM), we can skip the calculation of the ϕ part in Eq. (6) and directly find $k(x_i)\mathrm{k}(x_i{}')$, as shown in equation (7):

$$\mathcal{L}_{DS}(X, Y) = \| \frac{1}{n^2} \sum_{i,i'}^{n} k(x_i)k(x_i{}') - \frac{2}{mn} \sum_{i,j}^{n} k(x_i)k(y_j) + \frac{1}{m^2} \sum_{j,j'}^{n} k(y_j)k(y_j{}') \|_H \tag{7}$$

In the experiment, we use Gaussian kernels as the kernel function $k(\cdot)$. The larger the $\mathcal{L}_{DS}(X,Y)$ value, the more similar the two distributions; otherwise the smaller the $\mathcal{L}_{DS}(X,Y)$ value, the less similar the two distributions.

2) DVF SMOOTHNESS BASED REGISTRATION OPTIMIZATION

This paper calculates the displacement vector field smoothness as a deformation regularization (DR), which is shown in Eq. (8):

$$\mathcal{L}_{DR}(D_V) = \sum \|\partial x(D_V)\|_1 + \sum \|\partial y(D_V)\|_1 + \sum \|\partial z(D_V)\|_1 \tag{8}$$

where ∂x,∂y,∂z are the first-order partial derivatives of the displacement vector field D_V in the three directions of x, y and z, respectively.

In this paper, we use $\mathcal{L}_{DR}(D_V)$ to limit the degree of deformation of the moving PET image patches.

TABLE 1. Details about original PET/CT volume data.

Attribute	CT Volume	PET Volume
Size	512×512×(215-341)	128×128×(215-341)
Pixel spacing	$0.977 \times 0.977\, mm^2$	$5.469 \times 5.469\, mm^2$
Slice thickness	$3.270\, mm$	$3.270\, mm$

III. DATA AND PRE-PROCESSING

A. PRE-PROCESSING

For a given pair of a patient's original PET/CT volume, we pre-process the data through (a) calculate SUV and Hu value based on PET and CT volume, respectively. (b) down-sample the Hu images to the same resolution of the SUV images using bicubic interpolation. (c) limit the threshold to [-90, 300] for Hu, i.e. values that are smaller than -90 and larger than 300 are set to -90 and 300, respectively. Similarly, limit the threshold to [0, 5] for SUV. (d) standardize the data to have zero mean and unit variance, due to the large intensity differences among different volumes. (e) process each volume data with overlapping sampling to obtain training, validation and tset set.

B. DATA SET

To the best of our knowledge, there is no public dataset for whole-body PET/CT registration. The data used in this paper is acquired with a GE's PET/CT scanner (GE Discovery VCT) from the General Hospital of Shenyang Military Area Command, including whole-body PET and CT volumes of 176 patients in total. The PET scanner consists of a 32-ring detector using BGO crystals, while the CT scanner is a 64-slice spiral CT. Details about original PET/CT volume data are shown in Table 1. The size of original CT volume data is $512 \times 512 \times N$, while the corresponding PET volume data has the size of $128 \times 128 \times N$, where N is the number of slices of a patient. In this experiment, the value of N is between 215 and 341. All the PET/CT volumes are pre-processed with the method described in Section 3.A. To fit the limited 11GB GPU memory, the input of the DenseRegNet is sub-volumes with the size of $128 \times 128 \times 64$. Finally, 1159 pairs of SUV and Hu image patches with the size of $128 \times 128 \times 64$ are obtained.

In this paper, 663 pairs of SUV and Hu patches of 106 patients, 227 pairs of SUV and Hu patches of 35 patients and 259 pairs of SUV and Hu patches of 35 patients are randomly selected as training, validation and test set, respectively.

IV. EXPERIMENTS

A. EXPERIMENT SETTINGS

The DenseRegNet is implemented with Tensorflow. The number of convolutional layers in each DenseBlock is $C = 6$ with the kernel number of $G = 16$. Batch size is set to 4. Adam [28] optimizer is used in the experiment, and the learning rate is set to 0.01, $\beta_1 = 0.9$, $\beta_2 = 0.999$. 10-fold cross validation strategy is used in this experiment. We use six

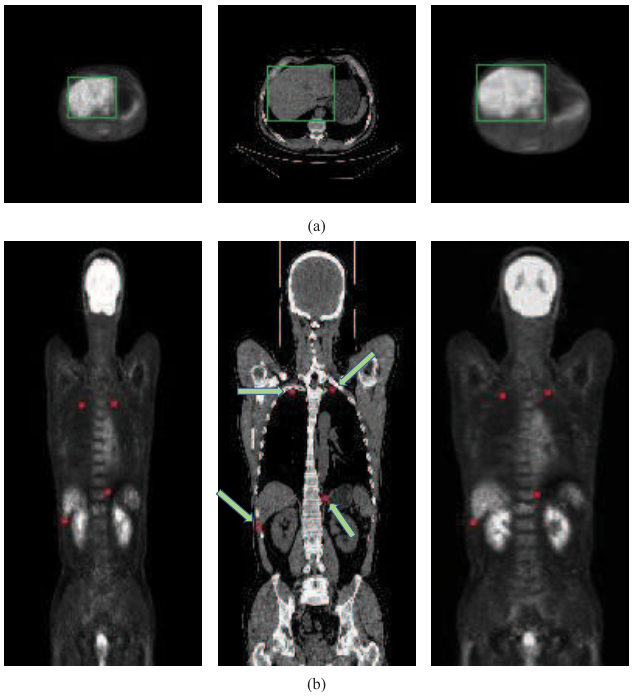


FIGURE 2. Visualization of liver bounding box and landmark points in both PET and CT used for calculating IoU and ED. The 1 - 3th column from left to right are corresponded to the source PET image, the source CT image and the deformed PET image. (a) Liver bounding box, shown in green, (b) Landmark points, shown in red.

Gaussian kernels with the sigma value of 2, 5, 10, 20, 40 and 80 to map the data to RKHS for calculating the distribution similarity.

All computations are performed on an Intel CoreTM ®i7-7700K CPU@3.60GHz, 8 Cores, 32 GB RAM running on Ubuntu 16.04 LTS and NVIDIA GeForce GTX 1080ti GPU with 3584 cores and 11GB memory.

B. EXPERIMENT RESULTS AND DISCUSSIONS

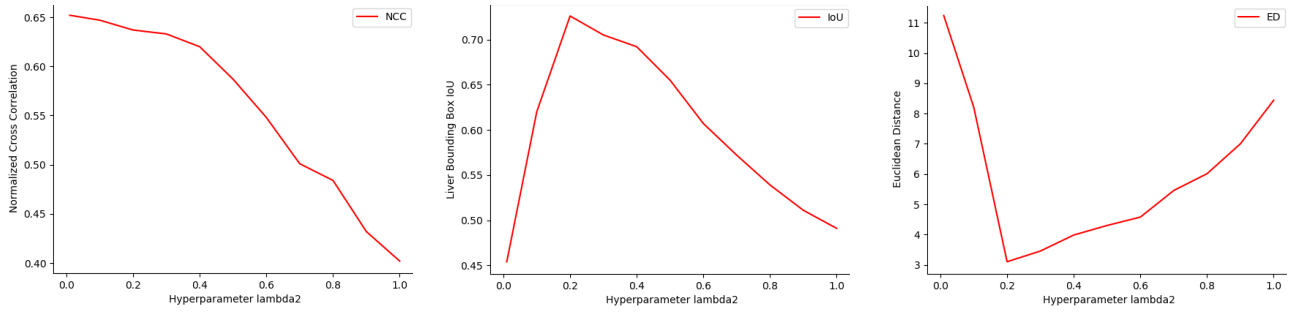
Up to date, there is no well-defined evaluation metric for deformable PET/CT image registration. Due to the natural characteristics of the PET/CT image registration task, it is not appropriate to use NCC as the sole evaluation metric. Liver bounding box IoU and ED on landmark points are additionally introduced as two metrics to evaluate registration results, anatomically. These two measures are defined as follows:

$$IoU(X, Y) = \frac{X_{Box} \cap Y_{Box}}{X_{Box} \cup Y_{Box}} \tag{9}$$

$$ED(FS_X, FS_Y) = \frac{1}{N} \sum_{i=1}^{N} d(point_i) \tag{10}$$

where X_{Box} and Y_{Box} are the liver bounding box of the input images X and Y. FS means the landmark points of one input image. $d(point_i)$ reprents the corresponding euclidean distance for each point of FS_X and FS_Y.

The liver bounding box and landmark points are obtained by manual-labeling PET and CT images before and after



(a) (b) (c)

FIGURE 3. Effect of varying the regularization parameter λ_2 on NCC, IoU and ED corresponds to subfigure (a), (b) and (c), respectively. The best results occur when $\lambda_2 = 0$ for NCC and $\lambda_2 = 0.2$ for IoU and ED.

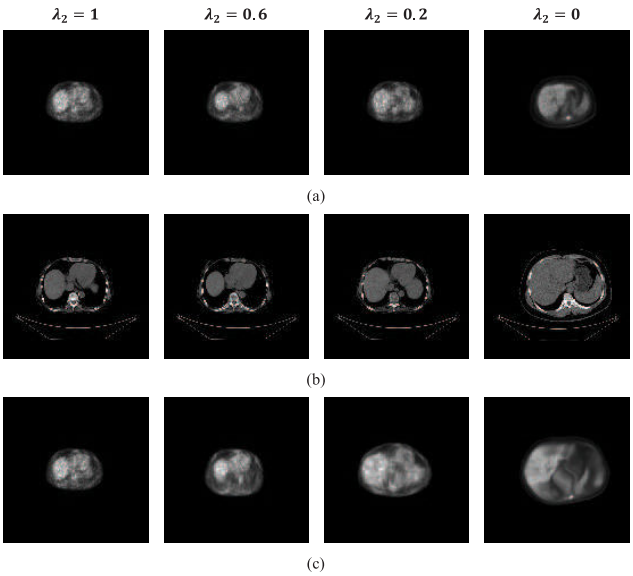


FIGURE 4. PET/CT images and the corresponding registration results with different λ_2 values. (a) PET images, (b) CT images, and (c) registration results. The 1 - 4th column from left to right are corresponded to the registration results with λ_2 values of 1, 0.6, 0.2, and 0, respectively.

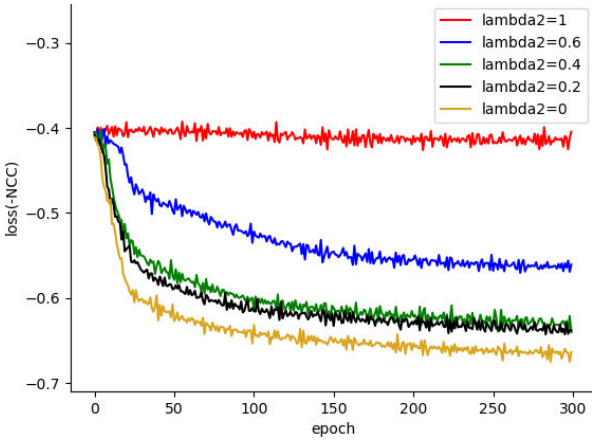


FIGURE 5. Learning curves showing the normalized cross correlation loss on the training set for different λ_2 values of DenseRegNet trained in 300 epochs for registration.

registration. The landmark points include two upper borders of the lung visible for both sides, the lower border of the right hepatic lobe and the left border of stomach, respectively. Figure2 visualizes the liver bounding box and landmark points in both PET and CT used for calculating IoU and ED.

1) SETTING OF SMOOTHING COEFFICIENT λ_2

The selection of deformation regularization coefficient λ_2 (i.e. the degree of deformation limitation by smoothing the displacement vector field) is critical for the results of PET/CT image registration [29]. In order to make a better comparison, we make qualitative and quantitative analyses of the registration results for different λ_2 values, maintaining the DenseRegNet network structure and other parameter settings unchanged, $\lambda_1 = 0$.

Figure 3 presents NCC, IoU and ED for the validation set for different values of the smoothing coefficient λ_2. The registration results with different values of the smoothing coefficient λ_2 are shown in Figure 4. Figure 5 shows $\mathcal{L}_{NCC}$ learning curves for different λ_2 values. Combined with the results in Figures 4, the deformation is totally limited when λ_2 takes the value of 1, thus there is almost no change before and after registration. And when λ_2 takes the value of 0.2, $\mathcal{L}_{NCC}$ converges to around 0.637, achieving the best registration results anatomically. Interestingly when setting $\lambda_2 = 0$, which enforces no regularization, results in the highest value of NCC. However, the PET image is excessively deformed caused by yielding displacement fields with some foldings and the worst values of IoU and ED are obtained with an unsmooth and unreasonable transformation. This is likely due to the fact that images of the two modalities, PET and CT, have large intensity differences in nature and a small number of differences in anatomical structure.

Through analysis, it is concluded that in the PET/CT image registration task, the variation tendency of NCC, IoU and ED are positively correlated within a certain degree of deformation. That is, a higher NCC value corresponds to a better IoU and ED, and also a better registration result. Due to the fact that the highlighted parts in CT images while having a low

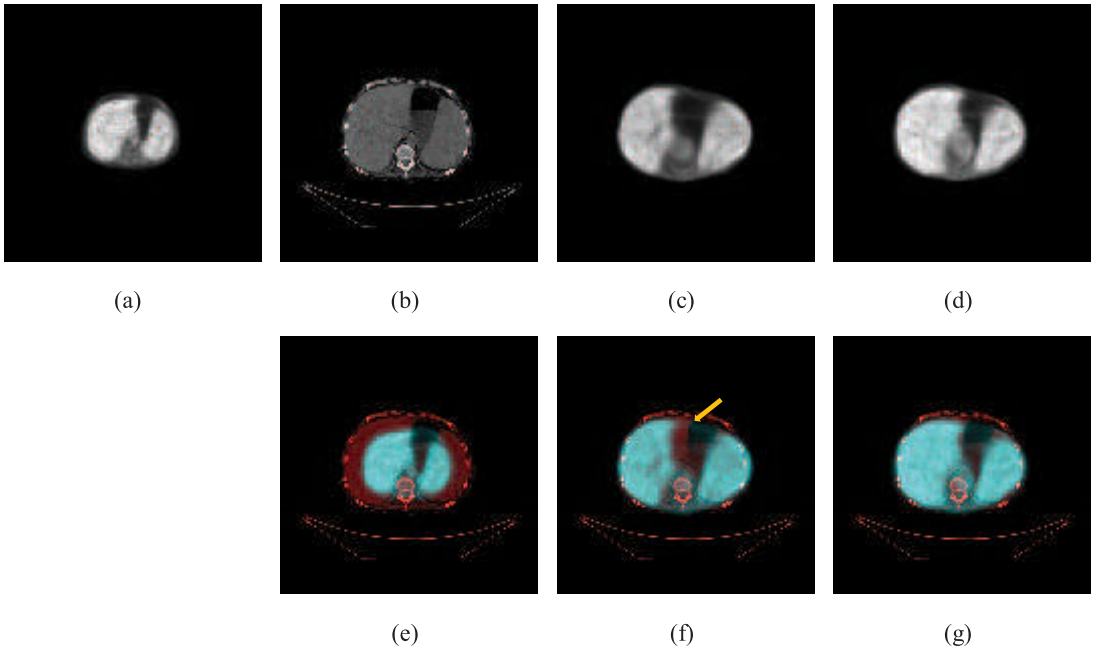


FIGURE 6. An example of PET/CT registration results with and without distribution similarity (DS) term. (a) The source PET image. (b) The source CT image. (c) The registration result without DS. (d) The registration result with DS. (e) The fusion image of the source PET and CT image. (f) The fusion image of the registered PET and CT image without DS. (g) The fusion image of the registered PET and CT image with DS.

TABLE 2. Quantitative evaluation of registration results with different λ_2 values.

λ_2	NCC	IoU	ED
1	0.407 ± 0.022	0.496 ± 0.015	8.051 ± 2.328
0.8	0.484 ± 0.065	0.539 ± 0.059	6.012 ± 1.599
0.6	0.548 ± 0.072	0.607 ± 0.070	4.577 ± 1.487
0.4	0.620 ± 0.066	0.692 ± 0.071	3.984 ± 1.369
0.2	0.637 ± 0.059	**0.726 ± 0.083**	**3.101 ± 1.502**
0	**0.651 ± 0.111**	0.556 ± 0.146	15.554 ± 3.716

contrast in PET images such as bones, when the deformation exceeds a certain degree, obtaining that a higher NCC value no longer indicates a better registration result. When beyond a certain degree of deformation (i.e. the value of λ_2 decreases from 0.2 to 0 in this experiment), the PET images have a tendency to be transformed into a ''mode-to-mode transition'' to the CT images, and the increase of NCC value will result in a worse registration result, instead. Therefore, we can draw a conclusion that IoU and ED are more appropriate in describing the anatomical similarity of the registration results than NCC.

The final quantitative evaluation results are discribed in Table 2.

By comprehensive analysis of NCC, IoU and ED with different λ_2 values for contrast experiments, we determine that the optimal value for smoothing coefficient λ_2 is 0.2.

2) EFFECTIVENESS OF DISTRIBUTION SIMILARITY

In order to further verify the effectiveness of the distribution similarity, we maintain the network structure and other parameter settings unchanged. Using grid search, λ_1 is set to 0.1, λ_2 is set to 0.2. The registration results with and without distribution similarity regularization are qualitatively and quantitatively analyzed. The registration results are shown in Figure 6 and the evaluation metrics are shown in Table 3 (DR and DS represent deformation regularization and distribution similarity term, respectively).

TABLE 3. Quantitative evaluation of the registration results with and without distribution similarity regularization.

Method	NCC	IoU	ED
NCC+DR	**0.637 ± 0.059**	0.726 ± 0.083	3.101 ± 1.502
NCC+DR+DS	0.633 ± 0.068	**0.740 ± 0.070**	**2.826 ± 1.495**

Qualitative and quantitative analyses show that the registration results are improved with the distribution similarity term. From Table 3, we can see that with the distribution similarity term, the value of NCC is a little decreased while IoU and ED are somehow improved. It can be seen from Figure 6 that there is a deviation in organ registration when using no distribution similarity during training. With the distribution similarity term, the data distribution between registered PET/CT images is more similar and the registration of organ contours is improved. The distribution similarity term acts to complement the measure of NCC and improve

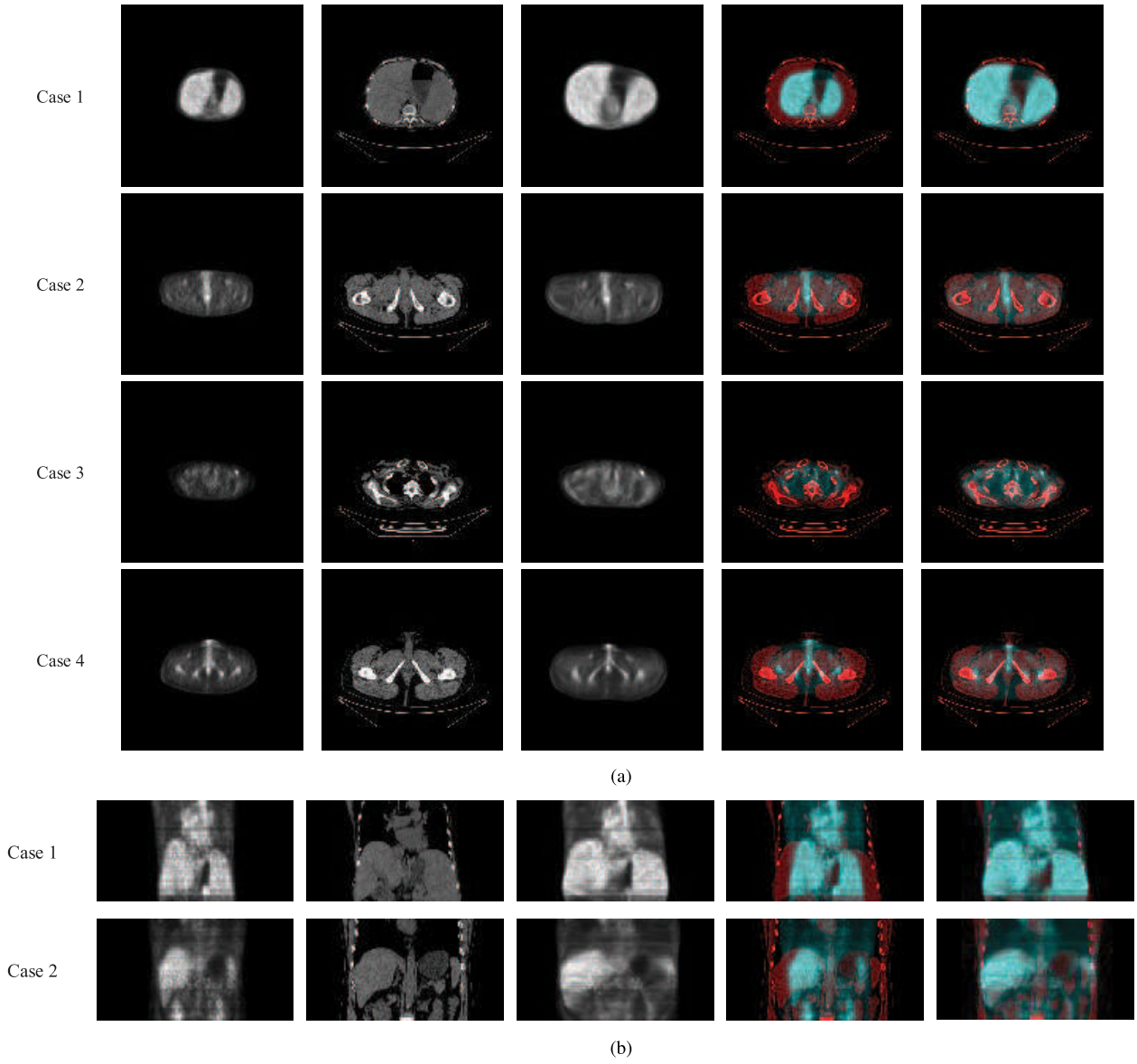


FIGURE 7. The final registration results with our proposed DenseRegNet. (a). An example of PET/CT registration results in transverse plane. (b). An example of PET/CT registration results in coronal plane. The 1 - 5th column from left to right are corresponded to the source PET image, the source CT image, the registration result, the fusion image of the source PET/CT image and the fusion image of the registered PET/CT image respectively.

the registration of the organ contour at the higher dimensional level.

The above results demonstrate the effectiveness of the distribution similarity term for deformable PET/CT registration.

3) COMPARISON WITH OTHER IMAGE REGISTRATION MODELS

We use the proposed DenseRegNet for 3D PET/CT image registration, maintaining $\lambda_1 = 0.1$ and $\lambda_2 = 0.2$. Registration results are obtained by adopting thg trained DenseRegNet to register the image patches in the test set using feedforward computation. The final registration results and fusion images of the proposed method in this paper are shown in Figure 7.

The quantitative evaluation for comparison with the existing registration models (Multi-thread ICP, DIRNet, FCN and VoxelMorph) is shown in Table 4 (DR and DS represent deformation regularization and distribution similarity, respectively).

It can be seen from Table 4 that our DenseRegNet results in a significant improvement over the other registration methods, achieving the best results in terms of liver bounding box IoU and ED on landmark points, and the second highest value of NCC. Generally speaking, the proposed method achieves improved performance results in the PET/CT registration task. A trained DenseRegNet can be used for registering a patient's volume data within only 10 seconds, which is much shorter than traditional registration methods.

TABLE 4. Quantitative evaluation for the test set for different registration methods.

Method	NCC	IoU	ED	Execution time(s/case)
Non-registration	0.402	0.434	11.241	N/A
Multithread-ICP [24]	0.632 ± 0.053	0.730 ± 0.086	3.028 ± 1.387	337.054 ± 45.303
2D DIRNet [18]	0.572 ± 0.062	0.591 ± 0.046	4.759 ± 2.489	2.775 ± 0.425
3D DIRNet [21]	0.573 ± 0.066	0.684 ± 0.061	3.611 ± 1.539	2.561 ± 0.541
FCN [19]	0.623 ± 0.051	0.711 ± 0.069	3.282 ± 0.961	3.017 ± 0.411
VoxelMorph [20]	**0.653 ± 0.109**	0.665 ± 0.097	4.994 ± 1.722	3.169 ± 0.706
NCC Only (Ours)	0.651 ± 0.111	0.556 ± 0.146	15.554 ± 3.716	**2.518 ± 0.576**
NCC+DR (Ours)	0.637 ± 0.059	0.726 ± 0.083	3.101 ± 1.502	2.832 ± 0.614
NCC+DR+DS (Ours)	0.633 ± 0.068	**0.740 ± 0.070**	**2.826 ± 1.495**	2.704 ± 0.593

V. CONCLUSION AND FUTURE WORK

We use deep learning methods for the deformable PET/CT registration task. We propose an unsupervised end-to-end image registration method, DenseRegNet, for effective and efficient 3D PET/CT registration. A two-level similarity measure together with a deformation regularization is used to optimize the training process: (a) NCC is used to measure the similarity of voxels at the global level. The degree of global matching can be shown by calculating the normalized cross correlation value between the moving image and the fixed image. (b) MMD is used to measure the similarity of data distributions at the higher dimensional level, which results in an organ-level registration improvement. (c) DVF smoothness is used as a deformation regularization to limit the degree of image deformation, which can improve the registration for those parts with large intensity differences between PET and CT images and obtain smooth and anatomically plausible spatial transformations. We use NCC, liver bounding box IoU and ED on landmark points to evaluate the registration results. Compared with other existing registration models [18]–[21], [24], qualitative and quantitative analyses show that our method achieves improved performance results in the challenging deformable PET/CT registration task. By introducing a new similarity measure and deformation regularization, this paper provides a new approach for registration between two modalities which have large intensity differences.

The method proposed in this paper has achieved certain effects in deformable PET/CT registration. It needs to be further verified whether the method can be usable for other multi-modality registration

Although CT is a structural image and PET is a functional image, the two still have some similarities in the anatomical structure. We believe that the effective use of this similarity is key achieve to better registration results. In our work, we implicitly improve the anatomical similarity between PET/CT images by limiting the displacement vector field and increasing the distribution similarity. In the future, we plan to develop a way to explicitly define the similarity of anatomical structures between PET/CT images for better registration results.

REFERENCES

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

[2] D. W. Townsend, J. P. Carney, J. T. Yap, and N. C. Hall, "PET/CT today and tomorrow," *J. Nucl. Med.*, vol. 45, pp. 4S–14S, Jan. 2004.

[3] A. Sotiras, C. Davatzikos, and N. Paragios, "Deformable medical image registration: A survey," *IEEE Trans. Med. Imag.*, vol. 32, no. 7, pp. 1153–1190, Jul. 2013.

[4] J. A. Schnabel, M. P. Heinrich, B. W. Papież, and J. M. Brady, "Advances and challenges in deformable image registration: From image fusion to complex motion modelling," *Med. Image Anal.*, vol. 33, pp. 145–148, Oct. 2016.

[5] B. Zitová and J. Flusser, "Image registration methods: A survey," *Image Vis. Comput.*, vol. 21, pp. 977–1000, Oct. 2003.

[6] J. Kybic and M. Unser, "Fast parametric elastic image registration," *IEEE Trans. Image Process.*, vol. 12, no. 11, pp. 1427–1442, Nov. 2003.

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

[8] T. Rohlfing, C. R. Maurer, D. A. Bluemke, and M. A. Jacobs, "Volume-preserving nonrigid registration of MR breast images using free-form deformation with an incompressibility constraint," *IEEE Trans. Med. Imag.*, vol. 22, no. 6, pp. 730–741, Jun. 2003.

[9] G. P. Penney, J. Weese, J. A. Little, P. Desmedt, D. L. G. Hill, and D. J. Hawkes, "A comparison of similarity measures for use in 2-D-3-D medical image registration," *IEEE Trans. Med. Imag.*, vol. 17, no. 4, pp. 586–595, Aug. 1998.

[10] T. D. Silva, A. Uneri, M. D. Ketcha, S. Reaungamornrat, G. Kleinszig, S. Vogt, N. Aygun, S.-F. Lo, J.-P. Wolinsky, and J. H. Siewerdsen, "3D–2D image registration for target localization in spine surgery: Investigation of similarity metrics providing robustness to content mismatch," *Phys. Med. Biol.*, vol. 61, no. 8, p. 3009, 2016.

[11] C.-R. Hwang, "Simulated annealing: Theory and applications," *Acta Appl. Math.*, vol. 12, no. 1, pp. 108–111, May 1988.

[12] K. Deb, A. Pratap, S. Agarwal, and T. Meyarivan, "A fast and elitist multiobjective genetic algorithm: NSGA-II," *IEEE Trans. Evol. Comput.*, vol. 6, no. 2, pp. 182–197, Apr. 2002.

[13] K. Lau and A. Chung, "A global optimization strategy for 3D-2D registration of vascular images," in *Proc. BMVC*, 2006, pp. 489–498.

[14] 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.

[15] 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. AAAI*, Feb. 2017, pp. 4168–4175.

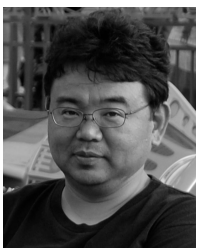
[16] H. Sokooti, B. de Vos, F. Berendsen, B. P. F. Lelieveldt, I. Išgum, and M. Staring, "Nonrigid image registration using multi-scale 3D convolutional neural networks," in *Medical Image Computing and Computer Assisted Intervention MICCAI*, M. Descoteaux, L. Maier-Hein, A. Franz, P. Jannin, D. L. Collins, and S. Duchesne, Eds. Cham, Switzerland: Springer, 2017, pp. 232–239.
[17] M. Jaderberg, K. Simonyan, A. Zisserman, and K. kavukcuoglu, "Spatial transformer networks," in *Proc. Adv. Neural Inf. Process. Syst.*, 2015, pp. 2017–2025.
[18] 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 *Proc. Int. Workshop Deep Learn. Med. Image Anal.*, 2017, pp. 204–212.
[19] H. Li and Y. Fan, "Non-rigid image registration using fully convolutional networks with deep self-supervision," Sep. 2017, *arXiv:1709.00799*. [Online]. Available: https://arxiv.org/abs/1709.00799
[20] G. Balakrishnan, A. Zhao, M. R. Sabuncu, J. Guttag, and A. V. Dalca, "An unsupervised learning model for deformable medical image registration," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2018, pp. 9252–9260.
[21] H. Yu, X. Zhou, H. Jiang, 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," *Proc. SPIE*, vol. 10953, Mar. 2019, Art. no. 109531X.
[22] 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.
[23] 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.
[24] Z. Song, H. Jiang, Q. Yang, Z. Wang, and G. Zhang, "A registration method based on contour point cloud for 3D whole-body PET and CT images," *BioMed Res. Int.*, vol. 2017, Feb. 2017, Art. no. 5380742.
[25] D. Khodadad, A. Ahmadian, M. Ay, A. F. Esfahani, H. Y. Banaem, and H. Zaidi, "B-spline based free form deformation thoracic non-rigid registration of ct and pet images," *Proc. SPIE*, vol. 8285, Sep. 2011, Art. no. 82851K.
[26] H. Yan, Y. Ding, P. Li, Q. Wang, Y. Xu, and W. Zuo, "Mind the class weight bias: Weighted maximum mean discrepancy for unsupervised domain adaptation," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2017, pp. 945–954.
[27] W. Zaremba, A. Gretton, and M. Blaschko, "B-test: A non-parametric, low variance kernel two-sample test," in *Proc. Adv. Neural Inf. Process. Syst.*, Jul. 2013, pp. 755–763.
[28] D. Kingma and J. Ba, "Adam: A method for stochastic optimization," in *Proc. Int. Conf. Learn. Represent.*, Dec. 2014.
[29] T. Rohlfing, "Image similarity and tissue overlaps as surrogates for image registration accuracy: Widely used but unreliable," *IEEE Trans. Med. Imag.*, vol. 31, no. 2, pp. 153–163, Feb. 2012.

HONGJIAN KANG is currently pursuing the M.S. degree with the Department of Software College, Northeastern University, China. His research interests include computer vision and medical image processing.

HUIYAN JIANG received the B.S. degree from the Department of Mathematics, Bohai University, China, in 1986, and the M.Sc. 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, and carried out the research of 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. 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 also 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 with Gifu University, Japan, where he is currently an Assistant Professor with the Department of Electrical, Electronic and Computer Engineering. His research interests include medical image analysis, medical image visualization, and pattern recognition.

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



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. He became a Research Assistant and an Associate Professor with the Faculty of Engineering, Gifu University, in 1995 and 2001, respectively, where he became an Associate Professor with the Department of Intelligent Image Information. He was a Visiting Associate Professor with the Department of Radiology, The University of Chicago, from 2008 to 2009. He became an Associate Professor with the, Gifu University, in 2017. He is currently a Professor with the Faculty of Engineering and the Chair of the Medical Research Department, Gifu University Artificial Intelligence Research Promotion Center, since 2019.

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. He became a Research Associate and an Associate Professor with the Gifu National College of Technology, in 1978 and 1986, respectively. He was a Visiting Researcher with the K. Rossmann Radiologic Image Laboratory, University of Chicago, from 1983 to 1986. He became an Associate Professor and a Professor with the Faculty of Engineering, Gifu University, in 1991 and 1995, respectively, where he has been a Professor and the Chair of intelligent image information with the Graduate School of Medicine, since 2002. He is currently a Research Professor with Gifu University and a Visiting Professor with Zhengzhou University, China.

YU-DONG YAO (F'11) received the B.Eng. and M.Eng. degrees from the Nanjing University of Posts and Telecommunications, Nanjing, China, in 1982 and 1985, respectively, and the Ph.D. degree from Southeast University, Nanjing, in 1988, all in electrical engineering. From 1987 to 1988, he was a Visiting Student with Carleton University, Ottawa, ON, Canada. From 1989 to 2000, he was with Carleton University, Spar Aerospace Ltd., Montreal, QC, Canada, and Qualcomm Inc., San Diego, CA, USA. Since 2000, he has been with the Stevens Institute of Technology, Hoboken, NJ, USA, where he is currently a Professor and the 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. He has 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 an Editor for the IEEE TRANSACTIONS ON WIRELESS COMMUNICATIONS, from 2001 to 2005. For his contributions to wireless communications systems, he was elected as a Fellow of the National Academy of Inventors, in 2015, and the Canadian Academy of Engineering, in 2017.