Region-Adaptive Deformable Registration of CT/MRI Pelvic Images via Learning-Based Image Synthesis

Xiaohuan Cao, Jianhua Yang, Yaozong Gao, Qian Wang, and Dinggang Shen, *Fellow, IEEE*

***Abstract*— Registration of pelvic computed tomography (CT) and magnetic resonance imaging (MRI) is highly desired as it can facilitate effective fusion of two modalities for prostate cancer radiation therapy, i.e., using CT for dose planning and MRI for accurate organ delineation. However, due to the large intermodality appearance gaps and the high shape/appearance variations of pelvic organs, the pelvic CT/MRI registration is highly challenging. In this paper, we propose a region-adaptive deformable registration method for multimodal pelvic image registration. Specifically, to handle the large appearance gaps, we first perform both CT-to-MRI and MRI-to-CT image synthesis by multi-target regression forest. Then, to use the complementary anatomical information in the two modalities for steering the registration, we select key points automatically from both modalities and use them together for guiding correspondence detection in the region-adaptive fashion. That is, we mainly use CT to establish correspondences for bone regions, and use MRI to establish correspondences for soft tissue regions. The number of key points is increased gradually during the registration, to hierarchically guide the symmetric estimation of the deformation fields. Experiments for both intra-subject and inter-subject deformable registration show improved performances compared with the state-of-the-art multimodal registration methods, which demonstrate the potentials of our method to be applied for the routine prostate cancer radiation therapy.**

***Index Terms*— Image synthesis, multi-modal registration, radiation therapy, learning-based registration.**

Manuscript received May 21, 2017; revised October 9, 2017; accepted March 14, 2018. Date of publication March 30, 2018; date of current version April 20, 2018. This work was supported in part by NIH under Grant AG053867 and Grant CA206100, in part by the National Key Research and Development Program of China under Grant 2017YFC0107600, in part by the National Natural Science Foundation of China under Grant 61473190, Grant 81471733, and Grant 61401271, and in part by the Science and Technology Commission of Shanghai Municipality under Grant 16511101100 and Grant 16410722400. The associate editor coordinating the review of this manuscript and approving it for publication was Prof. Oleg V. Michailovich. *(Corresponding authors: Qian Wang; Dinggang Shen.)*

X. Cao is with the School of Automation, Northwestern Polytechnical University, Xi'an 710072, China, and also with the Department of Radiology and BRIC, University of North Carolina at Chapel Hill, Chapel Hill, NC 27599 USA.

J. Yang is with the School of Automation, Northwestern Polytechnical University, Xi'an 710072, China.

Y. Gao is with Shanghai United Imaging Intelligence Company, Ltd., Shanghai 201807, China.

Q. Wang is with the Institute for Medical Imaging Technology, School of Biomedical Engineering, Shanghai Jiao Tong University, Shanghai 200030, China (e-mail: wang.qian@sjtu.edu.cn).

D. Shen is with the Department of Radiology and BRIC, University of North Carolina at Chapel Hill, Chapel Hill, NC 27599 USA, and also with the Department of Brain and Cognitive Engineering, Korea University, Seoul 02841, South Korea (e-mail: dgshen@med.unc.edu).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TIP.2018.2820424

I. Introduction

EXTERNAL beam radiation therapy (EBRT) [1] is an effective treatment for prostate cancer. In EBRT, computed tomography (CT) is of great importance for radiation dose planning. In order to largely avoid the side effects during EBRT, the beams should focus on prostate, and spare on the nearby normal tissues, i.e., both bladder and rectum. Thus, it is crucial to accurately delineate the main pelvic organs. However, the soft tissue contrast is low in CT as shown in Fig. 1, which makes it difficult to accurately contour the pelvic organs. To this end, magnetic resonance imaging (MRI) [2] is often used as a supplementary imaging modality in EBRT for better delineations of pelvic organs. As shown in Fig. 1, MRI has relatively high soft tissue contrast, thus making it easier to distinguish pelvic organs. To take the advantages of both CT and MRI for prostate cancer radiation therapy, it is important to accurately align CT and MRI, which often contain local deformations.

However, there are two main challenges for accurate and robust pelvic CT/MRI registration. **a) Huge appearance gap between CT and MRI.** For CT images, the bone regions are salient, but the contrast of soft tissues is usually low. While in MRI, it provides abundant appearance details especially for soft tissues. Therefore, it is difficult to define a simple metric that can guide the correspondence matching between the two modalities. **b) Large shape variations.** Pelvic organs are highly variable across individual subjects, as shown in Fig. 1. Even for the same patient, since CT and MRI are scanned at different devices, the shapes of organs may change dramatically due to inevitable rectum movement and possible bladder filling or emptying. Also, the prostate, bladder, and rectum are spatially close to each other. The shape changing of one organ can usually have a chain effect upon other neighboring organs.

A. Related Work

To tackle the challenges mentioned above, many methods have been proposed for multi-modal image registration, which can be generally divided into three categories: 1) *information-theory-based methods*, 2) *feature-based methods* and 3) *image-synthesis-based methods*.

1) Information-Theory-Based Registration Methods: The first way to solve multi-modal image registration is to find an effective similarity measurement, which can be used to detect correspondences between images of different modalities. Mutual information (MI) [3]–[5] is a typical method of this

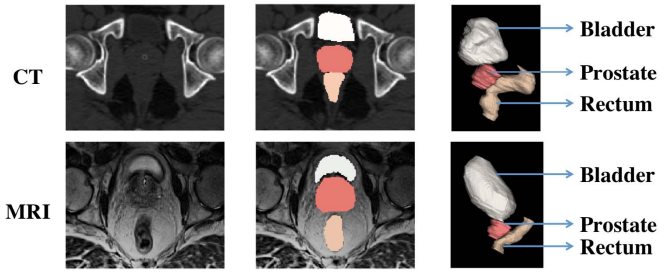


Fig. 1. Pelvic CT and MRI from two individual patients. The main pelvic organs, i.e., the prostate, bladder and rectum, are labeled in the images (middle) and also rendered (right).

category, which evaluates the multi-modal image similarity based on their (joint) histogram of intensities. Besides, other information-theory-based similarity metrics have also been introduced for multi-modal registration, including Kullback-Leibler divergence [6], Bhattacharyya divergence [7], Jensen-Shannon divergence [8], *etc*.

MI and its variants are widely applied [9]. However, it is limited in seeking local correspondences with image patches, as the size of image patches may not be large enough for precise histogram modeling. Meanwhile, it often ignores anatomical information, which is yet crucial for guiding correspondence detection and registration. To partially relieve this concern, various improved MI-based methods have been developed [10]–[16]. For example, the conditional mutual information (cMI) is proposed in [17] by incorporating the spatial dimension of location information. The spatial encoded mutual information (SEMI) incorporates a hierarchical weighting scheme to differentiate the contributions of the sample points in measuring the entropies [18]. Besides, the geometric and spatial context can also be considered when computing MI for high-resolution registration [19]. For multi-modal registration of eye fundus images, the feature neighborhood MI [20] is further proposed.

2) Feature-Based Registration Methods: Some feature-based registration methods are proposed to establish the correspondence by leveraging high-order appearance information, e.g., Gabor attributes [21], histogram of oriented gradient (HOG) [22], geometric moments [23]–[25], scale invariant feature transform (SIFT) [26], gradient location and orientation histograms (GLOH) [27], *etc*. For some kinds of features, they can be partially invariant to appearance differences across modalities when relying on gradient information. However, it is generally difficult to directly apply on multi-modal image registration, since the appearance gap across modalities can be high, e.g., between pelvic CT and MRI. A modality independent neighborhood descriptor (MIND) [28] is proposed based on the self-similarities around the neighborhood, in order to provide distinctive correspondences across different modalities. It can work well when the local anatomies and the structural patterns in small neighborhoods are similar. However, when the local anatomical variation becomes large, the robustness of MIND can be challenged.

3) Image-Synthesis-Based Registration Methods: Image synthesis has drawn a lot of interests and has been applied to many scenarios including image registration. To synthesize images across modalities, a model is often learned from a training dataset of pre-aligned multi-modal image pairs. When a new testing image comes, the image of the target modality can be synthesized via the learned model. So far, several learning-based methods have been developed for image synthesis. **a) Gaussian mixture regression model.** A CT image can be estimated from three MRI sequences by a Gaussian mixture regression model in [29], or from ultra-short echo-time MRI in [30]. In general, these methods are appropriate for mapping from "complex" image modalities (with rich appearance and anatomical details such as MRI) to "simple" image modalities (with limited anatomical details such as CT). But, the synthesized results are often fuzzy, and suffer from loss of many anatomical details. **b) Random forest regression model.** This method can synthesize high-resolution MRI from low-resolution acquisition with the help of other image modalities as demonstrated in [31]. However, the structured information is not used in this paper, which makes it less robust to preserve the anatomical information in the synthesized images. To address this, a structured random forest is applied in [32] to estimate CT from MRI, such that the neighboring anatomical information can be well-preserved in the synthesized images. **c) Deep learning model.** By using convolutional neural network (CNN), one can use MRI to synthesize PET [33] or CT [34]. However, large training datasets are often needed for deep-learning-based image synthesis methods.

By applying image synthesis, one can convert the difficult multi-modal registration to the relatively easy mono-modal registration problem. The main idea is that the large appearance gap between different modalities can be bridged by synthesizing one modality from another modality. In this way, the conventional deformable registration methods can be directly applied to registering two similar-looking images. The image synthesis has been applied to the registration of ultrasound image with MRI [35] or with CT [36]. Similarly, a solution to CT/MRI registration can be found in [37].

However, there are still three main limitations related to the image-synthesis-based multi-modal registration methods. **a) Image synthesis is always performed in a single direction only.** Thus, this single-directional image synthesis introduces bias when used to guide registration, as only one modality contributes its anatomical information to guide the correspondence detection. **b) Most existing studies focus on rigid/affine registration by ignoring the more challenging deformable registration problem.** In the literature, many previous works synthesize "simple" modal images (e.g., CT) from the "complex" modal images (e.g., MRI). The loss of rich anatomical information in the "simple" modality makes it hard to conduct accurate deformable registration. To use those abundant anatomical details for steering the accurate deformable registration, the "simple" to "complex" image synthesis is also necessary. However, although the "complex" image modality (e.g., MRI) can sometimes be synthesized, it often requires the input of multiple image modalities, which

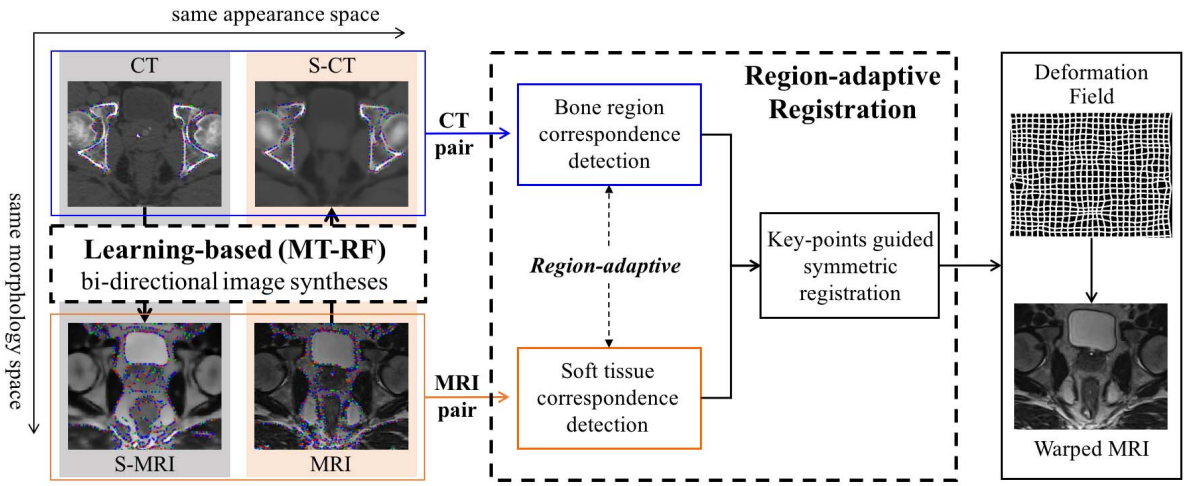


Fig. 2. Method overview: the proposed region-adaptive deformable registration method for multi-modal pelvic images. There are two main components in our method as indicated by dashed box: a) Learning based bi-directional image synthesis using Multi-Target Regression Forest (MT-RF), and b) Region-adaptive Registration. S-CT and S-MRI denote the synthesized CT and the synthesized MRI, respectively.

limits its practical application. **c) The existing studies mainly consider intra-subject multi-modal registration.** For the case of inter-subject pelvic image registration, it is more challenging due to large shape/appearance variation across subjects.

B. Contributions

Our previous work [38], [39], which is the first of using bi-directional image synthesis for multi-modal pelvic image registration, achieves promising performance for *intra-subject* registration. This bi-directional image synthesis framework provides a new solution for multi-modal registration by taking advantage of the complementary information from both modalities. However, there are still two issues that need to be further addressed. *First*, the performance of image synthesis plays an important role for the subsequent registration, which should be steered by accurate anatomical correspondence. *Second*, it is still challenging to effectively integrate the complementary information from both modalities for accurate and robust registration. To address above issues, in this paper, we propose a novel *region-adaptive* registration method to better tackle the intra- and inter-subject multi-modal pelvic image registration problem. Our main contributions can be summarized as follows.

1) Multi-Target Regression Forest (MT-RF): The MT-RF model is proposed for more accurate bi-directional image synthesis, i.e., synthesizing CT from MRI and also synthesizing MRI from CT. Compared with the previous single-target regression forest (ST-RF), the novel MT-RF model adopts both CT and MRI modalities for joint supervision of the learning for directional image synthesis. We show in experiments that MT-RF can bridge the appearance gap between the two modalities and produce more accurate synthesis results. Thus, we are able to acquire more discriminative anatomical details from the synthesized images to steer the deformable registration of multi-modal pelvic images more effectively.

2) Region-Adaptive Registration: In order to account for large local deformations among pelvic organs, a novel region-adaptive registration framework is proposed to guide accurate and robust local correspondence matching during deformable registration. Instead of treating the CT and MRI modalities equally across the entire image space, the distinctive anatomical details of each modality are specially treated for individual anatomical regions, and then combined for joint estimation of the dense deformation fields. In this way, our multi-modal image registration can benefit from the distinctive anatomical details provided by each individual image modality, where large local deformations among the main pelvic organs can be more effectively estimated.

3) Intra- and Inter-Subject Registration: In order to fully evaluate the effectiveness and flexibility of our method, we conduct comprehensive experiments on both *intra-subject* and more challenging *inter-subject* registration tasks. The experimental results show that the proposed method can successfully solve the *inter-subject* pelvic image registration problem, with promising accuracy and robustness.

II. Method

A. Overview

Fig. 2 shows the flowchart of our proposed region-adaptive deformable registration method based on bi-directional image synthesis. **First**, a new learning based method, *multi-target regression forest* (MT-RF) is proposed to learn the appearance mapping in bi-directions, i.e., from MRI to CT and also from CT to MRI, in order to eliminate the appearance difference across modalities and meanwhile preserve abundant anatomical details. Specifically, two image pairs can be obtained after image synthesis, i.e., a) the CT pair that includes an actual CT and a synthesized CT (S-CT), and b) the MRI pair that includes an actual MRI and a synthesized MRI (S-MRI), as shown in Fig. 2. The details of the MT-RF for bi-directional image synthesis will be described in Section II.B. **Then**,

to fully use the complementary information from both modalities, we propose the *region-adaptive registration* approach by detecting a) bone correspondences from the CT pair and b) soft tissue correspondences from the MRI pair. In this way, the abundant anatomical details in both modalities can be combined effectively to steer the multi-modal image registration and jointly estimate the deformation field. Moreover, we also derive the *symmetric* registration strategies, which is effective in dealing with large local deformations of pelvic organs. The details of region-adaptive registration will be described in Section II.C.

B. Learning-Based Bi-Directional Image Synthesis

The quality of the synthesized images will highly influence the registration performance. Thus, it is necessary to build an accurate and robust mapping between CT and MRI. To this end, we propose a new *multi-target regression forest* (**MT-RF**). As it is more difficult to synthesize MRI from CT, we use the CT-to-MRI mapping as an example in this paper. Note that the same method can be easily transferred for the MRI-to-CT mapping, although slightly different settings may be adopted (c.f. the detailed settings provided in Sec III.B).

1) Multi-Target Regression Forest (MT-RF): Random forest is a popular machine learning technique, which can be used for classification and regression [40]–[43] with high training and testing efficiency. The detailed introduction of the general random forest can be found in [40] and [44]. For image synthesis, a random forest regression model is learned to map one modality to the other modality. Intuitively, for CT-to-MRI synthesis, the MRI intensity patch can be estimated from the features extracted from the corresponding CT patch [39]. In the setting of the single-target regression forest (ST-RF), only one target, i.e., the MRI intensity patch, is used to supervise the training process. The mapping could be difficult to learn, especially when the target MRI intensity patch (e.g., with relatively smooth appearance) cannot provide distinctive and salient appearance information to supervise the training. Therefore, we propose the *multi-target regression forest* (MT-RF), where both CT and MRI intensity patches are used as multiple targets, which can jointly supervise the growth of each tree. In this way, the abundant anatomical details from both modalities can be utilized comprehensively for more accurate and efficient modeling of image synthesis.

a) Sampling: For the CT-to-MRI synthesis, as shown in Fig. 3, the training stage requires a CT/MRI paired dataset where different training subjects have been pre-aligned (the processing details are explained in the Experiment section). For preparation of training, M pairs of training samples are drawn randomly from the corresponding locations of the two image modalities of the same subject in the training dataset. The input to the MT-RF includes the features of the training samples extracted from the source CT domain. The output of MT-RF includes two (regression) targets: a) **Target 1** is the patch intensity of the **target MRI,** and b) **Target 2** is the patch intensity of the **source CT** from which the input features are extracted. Compared with the case of using just a single target in ST-RF (i.e., using only Target 1), including the source CT as an additional target contributes to better modeling of the mapping from the features of CT to both MRI and CT patches, which will be demonstrated in the experiments.

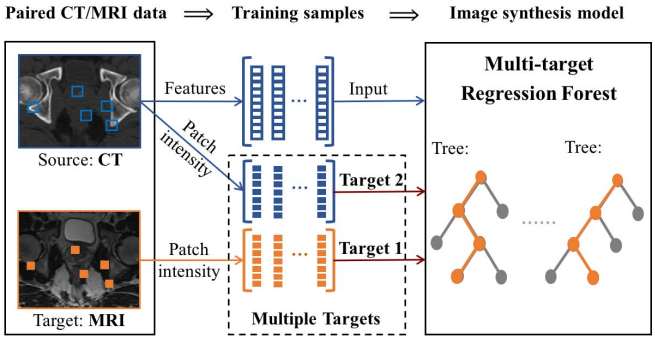


Fig. 3. Illustration of the training stage of CT-to-MRI synthesis by multi-target regression forest (MT-RF).

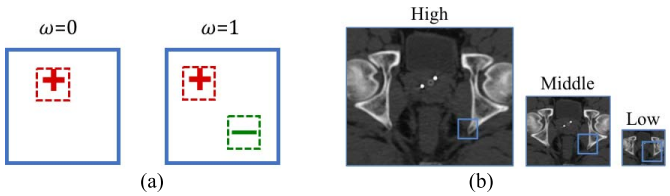


Fig. 4. Illustration of the extraction of the Haar-like features. In (a), we compute each Haar-like feature as the mean intensity value of a randomly generated "positive" block, or the difference of mean intensities of the "positive" and the "negative" blocks. In (b), we extract the Haar-like features from multi-resolution images for better representation of the sampled location in the center of the patch.

b) Feature extraction: The features used to represent each sample voxel are extracted from its local patch in the source CT domain, and they are essential to learn an effective CT-to-MRI mapping. In this work, we adopt the Haar-like features [32], [45] to describe the low-level image appearances, which are powerful and can be extracted efficiently [46]. Specifically, two types of the Haar-like features are employed: a) the mean intensity value within a "positive" block, and b) the difference of the mean intensities of a "positive" and a "negative" block, as shown in Fig. 4(a). Mathematically, each Haar-like feature can be defined as:

$$f[p(x)|\mu_+, r_+, \mu_-, r_-, \omega] = \frac{\sum b_{r_+}(\mu_+)}{(2r_+ + 1)^3} - \frac{\omega \cdot \sum b_{r_-}(\mu_-)}{(2r_- + 1)^3}, \tag{1}$$

where $p(x)$ is the local patch centered at x. $f[\cdot]$ denotes a certain Haar-like feature computed with the parameters $\{\mu_+, r_+, \mu_-, r_-, \omega\}$. Specifically, μ_+, r_+ are the center and radius of the "positive" block, and μ_-, r_- for the "negative" block. $\omega \epsilon \{0, 1\}$ represents the type of Haar-like features, i.e., $\omega = 0$ for the first type, and $\omega = 1$ for the second type.

Moreover, we extract the Haar-like features in the low, middle and high image resolutions by considering multi-resolution appearance information for better representation of each sample voxel, as shown in Fig. 4(b). The original (high-resolution) image is first down-sampled to obtain the middle- and low-resolution images. Then, we fix the patch size

(in voxel) to extract patches from three resolution images at the corresponding location. The parameters $\{\mu_+, r_+, \mu_-, r_-, \omega\}$ are randomly generated for the extraction of abundant Haar-like features for the patches. Finally, we concatenate all the extracted features into the final feature vector for the sampled location. Note that, zero padding is applied when the extracted patches exceed the image size.

c) Training: Random forest consists of multiple decision trees, and each tree is trained independently. The input to train each tree is the feature vectors $\boldsymbol{F} = [\boldsymbol{f}_1, \boldsymbol{f}_2, \ldots, \boldsymbol{f}_M]$ along with the two corresponding targets $T^1 = [t_1^1, t_2^1, \ldots, t_M^1]$ and $T^2 = [t_1^2, t_2^2, \ldots, t_M^2]$. Here, M is the number of the samples. To adapt for our MT-RF, we define the following objective function:

$$\arg\max_{\boldsymbol{f}^k, \tau} \lambda G_1 + (1-\lambda) G_2, \quad 0 < \lambda < 1, \tag{2}$$

$$G_i = \frac{1}{\sigma_i}\left(V_i(N) - \frac{|N^L|}{|N|} V_i(N^L) - \frac{|N^R|}{|N|} V_i(N^R)\right), \quad i = 1, 2, \tag{3}$$

$$N^L = \{(\boldsymbol{f}, t^1, t^2) \in N | \boldsymbol{f}^k < \tau\},$$
$$N^R = \{(\boldsymbol{f}, t^1, t^2) \in N | \boldsymbol{f}^k > \tau\}. \tag{4}$$

For each node splitting, $\boldsymbol{f}$ is the subset of the features that are extracted from the sampled patches of the source CT. $\boldsymbol{f}^k$ is the k-th feature that is chosen from the feature vector for node splitting, and τ is the feature threshold that separates the samples into the left and the right sub-nodes (corresponding to N^L and N^R). G_i is the information gain for the regression target i, and λ is the weight to balance the two targets in MT-RF. The growth of each tree aims to find the optimal $\{\boldsymbol{f}^k, \tau\}$ to maximize the combined information gain when splitting the node into the left and right sub-nodes. $V_i(\cdot)$ encodes the variance of the training samples regarding the regression target i. Since we use the patch intensity as the regression target, the variance here computes the mean value of all element-wise variances in the target. $|\cdot|$ is the number of the samples in each node. Accordingly, $V_i(N^L)$ and $V_i(N^R)$ represent the variances of the training samples in the left and right sub-nodes. Note that, since the intensity distribution/range is quite different in CT and MRI, the normalization term σ_i is introduced to account for the average variance at the root node for each regression target.

d) Testing: When a new CT image comes in the testing stage, each voxel is regarded as a sample, from which the respective features are extracted. Then, these features are fed to the trees in the trained MT-RF model. Next, the average of the outcomes of all trees is used as the synthesized MRI patch for the testing CT sample, by excluding the CT domain information in the output. Since we perform patch-wise synthesis by traversing all the voxels in the new CT image, the estimated patches are highly overlapping. The final intensity of each voxel can be obtained by averaging all the overlapping estimates. Then, the whole synthesized MR image can be obtained eventually.

2) Auto-Context Refinement: The appearance mapping between CT and MRI is complex and highly non-linear. In order to further refine the above one-shot mapping by MT-RF, the auto-context model [47] is leveraged to improve the quality of the synthesized image iteratively, as described below.

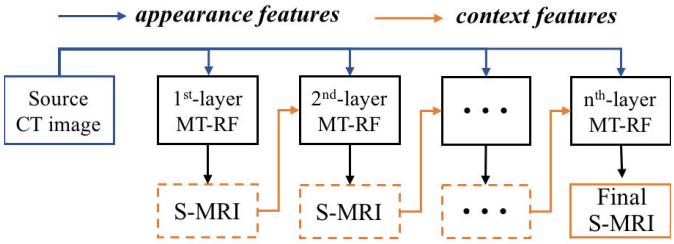


Fig. 5. Illustration of the auto-context model for multi-layer MT-RF that performs CT-to-MRI synthesis.

Specifically, for synthesizing MRI from CT, after the 1st-layer MT-RF (as shown in Fig. 5), a tentative S-MRI can be obtained. Then, the same Haar-like features (namely *context features*) are also extracted from the tentative S-MRI. The 2nd-layer MT-RF can then be built by the new feature vector, which concatenates the *context features* with the original *appearance features*. In particular, the *appearance features* are the Haar-like features extracted from the source CT, while the *context features* are also the Haar-like features but extracted from the tentative S-MRI. After that, the S-MRI can be updated as the output of the 2nd-layer MT-RF. As shown in Fig. 5, the same procedure can be performed iteratively, which results in a chain of cascaded MT-RFs. In this auto-context manner, the synthesized MRI can be improved layer-by-layer, by incorporating both the updated context features and the original appearance features.

C. Region-Adaptive Symmetric Registration

After bi-directional image synthesis, the appearance differences are largely eliminated in the CT pair and the MRI pair. Meanwhile, all the anatomical details are preserved in both modality pairs. Obviously, CT has high contrast in bone regions, while MRI has rich anatomical details in soft tissues. To accurately estimate the large local deformations between pelvic CT and MRI, we further propose the *region-adaptive* registration method, for combining the complementary anatomical details in both modalities to detect the correspondences for key points. Specifically, spline-based deformable model is adopted to estimate the deformation field between CT and MRI *hierarchically* and *symmetrically,* based on the key points correspondences.

Generally, the conventional objective function for direct registration of the original CT and MRI can be defined as:

$$\phi = \arg\min_{\phi} \mathcal{M}\left(I_{\mathrm{CT}}, \mathcal{D}\left(I_{\mathrm{MR}}, \phi\right)\right) + \alpha \mathcal{R}(\phi), \tag{5}$$

where I_{CT} and I_{MR} are the original CT and MRI, and $\mathcal{M}$ is a metric used to measure the dissimilarity between the CT and MRI. $\mathcal{D}$ is an operator used to warp the subject MRI via the estimated deformation field ϕ. $\mathcal{R}$ is a regularization term to constrain the smoothness of ϕ. This objective function aims to optimize the deformation field ϕ by minimizing

the discrepancy between CT and the warped MRI. After bi-directional image synthesis, the CT pair (i.e., CT I_{CT} and S-CT $I_{\mathrm{S-CT}}$) and the MRI pair (i.e., MRI I_{MR} and S-MRI $I_{\mathrm{S-MR}}$) are obtained to guide the registration. Then, the objective function in Eq. (5) can be revised as follows:

$$\phi = \arg\min_{\phi} \mathcal{M}\left(I_{\mathrm{CT}}, \mathcal{D}\left(I_{\mathrm{S-CT}}, \phi\right)\right) + \mathcal{M}\left(I_{\mathrm{S-MR}}, \mathcal{D}\left(I_{\mathrm{MR}}, \phi\right)\right) + \alpha\mathcal{R}(\phi). \tag{6}$$

That is, we optimize the deformation field ϕ by simultaneously minimizing the discrepancy within both the CT pair and the MRI pair. In this way, we avoid evaluating image similarities across modalities, while the anatomical details of both modalities are incorporated for registration.

1) Key Points: To effectively solve Eq. (6), we sample **key points** to guide the registration [48]. The key points are sampled in the regions with rich anatomical information, for the sake of establishing accurate and robust local correspondences. Mathematically, the key points can be sampled by following the importance:

$$\mathcal{P} = \frac{|\nabla_x| + |\nabla_y| + |\nabla_z|}{|\nabla_G|}, \tag{7}$$

where ∇_x, ∇_y, and ∇_z are image gradient operators in three directions, and $|\cdot|$ computes the magnitude. $|\nabla_G|$ is the maximum gradient magnitude for each individual image to normalize $\mathcal{P}$ to [0,1]. Obviously, the key points are distributed in the whole image volume, while the density tends to be high on the informative regions (e.g., organ boundaries) and low in the smooth regions (e.g., inside the bladder or bones). For each key point, the local patch is extracted as its *morphological signature* to guide the correspondence detection. In each modality pair, particularly, we use the normalized cross-correlation (NCC) as the metric to search for the correspondence for the key point. The locations with high NCC values are regarded as the candidates for correspondence matching.

2) Region-Adaptive Selection of Key Points: Apparently, CT has rich anatomical details in bone regions while MRI has rich anatomical details in soft tissue regions. The *region-adaptive* registration can thus use all the anatomical details in both CT and MRI, and retain the distinctiveness of the informative regions in each modality for the estimation of the deformation field. Fig. 6 illustrates the necessity of the proposed region-adaptive registration method. The similarity map shows the NCC values calculated from three key points in the template space (i.e., CT or S-MRI) to the candidates in the subject space (i.e., S-CT or MRI). For the CT pair, we observe good correspondence for the first key point that is located in the bone region. For the MRI pair, the similarity maps prompt well-established correspondences for the second and the third key points, both of which belong to soft tissues.

Specifically, the key points of bone regions are sampled from the CT pair by truncating the importance probability in Eq. (7) with a threshold. For the MRI pair, bone regions are temporarily suppressed by a bone mask generated from the corresponding S-CT image, thus the key points of soft tissue regions can be sampled from the MRI pair only. Next,

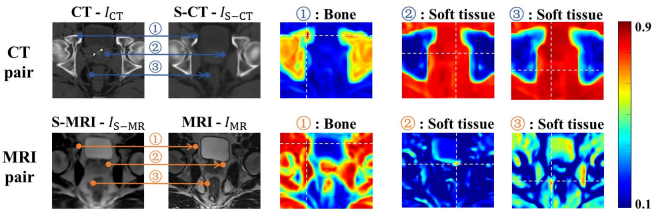


Fig. 6. Similarity maps for bone and soft tissue regions obtained from CT pair (upper row) and MRI pair (bottom row), respectively. All the NCC values are normalized to [0,1].

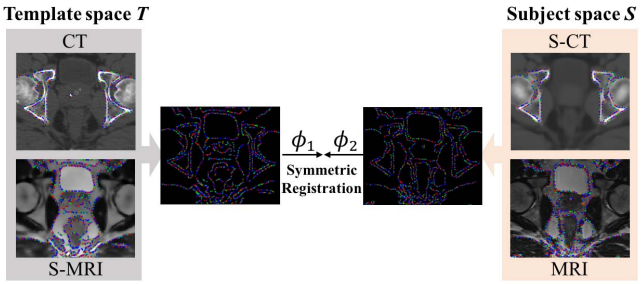


Fig. 7. Illustration of our proposed region-adaptive registration method implemented in the hierarchical and symmetric manner. Blue, red and green points indicate the key points selected in three different hierarchies.

the morphological signature for each key point is extracted from its respective modality. As shown in Fig. 7, the key points cover the whole image space eventually.

3) Hierarchical and Symmetric Implementation: For the robustness of the multi-modal pelvic image registration, we exploit the symmetric diffeomorphic scheme for the proposed region-adaptive registration and preserve the inverse-consistency [49]–[51] of the estimated deformation field. Specifically, the key points selected from both the template space and the subject space guide the images to warp simultaneously to an intermediate space, which results in two deformation fields: a) ϕ_1 from the template space to the intermediate space and b) ϕ_2 from the subject space to the intermediate space, as shown in Fig. 7.

Mathematically, based on Eq. (6), the objective function of our **region-adaptive symmetric** registration can be modified as follows:

$$\{\phi_1, \phi_2\} = \arg\min_{\phi_1,\phi_2} \mathcal{M}\left(\mathcal{D}\left(T, \phi_1\right), \mathcal{D}\left(S, \phi_2\right)\right) + \alpha\left(\mathcal{R}(\phi_1) + \mathcal{R}(\phi_2)\right), \tag{8}$$

where T and S represent the template and the subject spaces. The CT pair (I_{CT} and $I_{\mathrm{S-CT}}$) is used to guide the bone region registration, and the MRI pair ($I_{\mathrm{S-MR}}$ and I_{MR}) is used to guide the soft tissue registration. Note that the template and the subject spaces can be arbitrarily exchanged in Eq. (8). Here, the metric $\mathcal{M}$ computes the overall distance between the correspondence points across the two (warped) images, while $\mathcal{D}$ is the warping operation to warp the template T and the subject S under the deformation field ϕ_1 and ϕ_2, respectively. $\mathcal{R}$ is the bending energy following the thin-plate spline (TPS) based deformable model [52], [53].

To make the registration more robust and accurate, we further embed the above procedure into a hierarchical framework.

Algorithm 1 Region-Adaptive Key-Points Guided Deformable Registration

Data: CT: I_{CT}; synthesized CT: $I_{\mathrm{S-CT}}$; MRI: I_{MR}; synthesized MRI: $I_{\mathrm{S-MR}}$
Result: Forward deformation field: ϕ; Backward deformation field: ϕ^{-1}
Initialization: $i = 0$;
Template space T: $I^0_{\mathrm{CT}} = I_{\mathrm{CT}}$, $I^0_{\mathrm{S-MR}} = I_{\mathrm{S-MR}}$;
Subject space S: $I^0_{\mathrm{S-CT}} = I_{\mathrm{S-CT}}$, $I^0_{\mathrm{MR}} = I_{\mathrm{MR}}$;
ϕ_1=Identity, ϕ_2=Identity, as shown in Fig. 7;
Incremental deformation: φ^0_1 = Identity, φ^0_2 = Identity.

While (i< MaxIterationNumber)
- $i = i + 1$;
- Warp the template images: $I^i_{\mathrm{CT}} = \mathcal{D}(I^{i-1}_{\mathrm{CT}}, \phi_1)$, $I^i_{\mathrm{S-MR}} = \mathcal{D}(I^{i-1}_{\mathrm{S-MR}}, \phi_1)$.
- Warp the subject images: $I^i_{\mathrm{S-CT}} = \mathcal{D}(I^{i-1}_{\mathrm{S-CT}}, \phi_2)$, $I^i_{\mathrm{MR}} = \mathcal{D}(I^{i-1}_{\mathrm{MR}}, \phi_2)$.
- Region-adaptive key-points selection:
  - Bone region based on $[I^i_{\mathrm{CT}}, I^i_{\mathrm{S-CT}}]$;
  - Soft tissue based on $[I^i_{\mathrm{S-MR}}, I^i_{\mathrm{MR}}]$.
- Obtain the incremental deformation based on selected key points via local correspondence matching.
- Obtain the incremental dense deformation field φ^i_1, φ^i_2.
- Update the estimated deformation field:
  - $\phi_1 = \phi_1 \circ \varphi^i_1$;
  - $\phi_2 = \phi_2 \circ \varphi^i_2$

End While

Repeat the above procedures in three levels: low, middle and high image resolution. For each level, the deformation field is initialized by the final result obtained by the previous image resolution level.

1: For key points selection, the key points number is gradually increased for every iteration.
2: For local correspondence matching, the patch size is gradually decreased for every iteration.

Initially, only a small set of key points with very high importance are selected to drive the registration, as indicated by the red points in Fig. 7. These key points are distributed non-uniformly, yet cover the whole image volume. Obviously, the density of the key points tends to be high in regions with rich anatomical details, which is crucial to robust and accurate correspondence detection. These key points contribute their detected correspondences to the estimation of the dense deformation fields obtained by TPS [53] interpolation. As the template and the subject move closer to each other, more key points are gradually added. Note that, during the hierarchical matching, the scale of the image patch, which is used as the morphological signature of each key point, decreases hierarchically. In this way, the key points with global distinctiveness are utilized in the initial stage to drive the robust registration, while the deformation field is refined via more key points and their abundant local appearance details. The detailed implementation is summarized in **Algorithm 1.**

For every iteration, i.e., hierarchy, we estimate the incremental deformation field, which is composed to the diffeomorphic deformation field as in [51]. The final end-to-end deformation ϕ can be obtained by composing the two deformation fields ϕ_1 and ϕ_2:

$$\begin{aligned} \text{Forward deformation: } \phi &= \phi_2 \circ \phi_1^{-1}, \\ \text{Backward deformation: } \phi^{-1} &= \phi_1 \circ \phi_2^{-1}. \end{aligned} \tag{9}$$

Here, "$\circ$" means deformation composition [54], ϕ indicates the forward deformation field, and ϕ^{-1} is the inverse of ϕ.

III. Experimental Results

A. Data Description and Preprocessing

The experiment dataset includes 20 prostate cancer patients, each of which has a pair of CT and MR pelvic images. The original CT images were acquired from a Philips scanner with the image size of 512×512×(443∼509) and the resolution of 1.17×1.17×1mm^3. The original MR images were scanned from a Siemens Avanto scanner with the image size of 256×256×(144∼160) and resolution of 1×1×1 mm^3. The prostate, bladder and rectum were manually delineated by an experienced radiation oncologist, which serve as the ground-truth in the registration experiments below.

For data preprocessing, all the MRI data are first performed with N3 bias correction to eliminate the intensity inhomogeneity [55]. Then, we linearly align the CT and MRI of the same patient by using MI as the similarity metric [56]. Next, both CT and MRI are cropped to the same size of 200×200×80, by removing the unrelated regions. Note that the cropped images are large enough to include the prostate, bladder and rectum regions.

For the learning-based image synthesis, a *pre-aligned dataset* is needed for training. To guarantee the high precision in multi-modal pre-alignment, here we use the manual labels of the main pelvic organs to guide accurate intra-subject registration for preparing the training data. Specifically, for the linearly aligned CT and MRI of the same patient (using FLIRT [56]), we first adopt SyN [49] to complete their deformable registration based on their intensity images, with careful parameter tuning. Then, we apply diffeomorphic Demons [54] to refine the registration by aligning the manual labels of the prostate, bladder and rectum. Finally, affine registration is applied to roughly register all subjects to a common space. Note that all the transformations are combined together, such that a certain image only needs to be warped for once.

B. Experiments on Learning-Based Image Synthesis

In this section, we use the *pre-aligned dataset* to perform the experiments on a) synthesizing MRI from CT (**CT-to-MRI mapping**) and b) synthesizing CT from MRI (**MRI-to-CT mapping**), based on our proposed multi-target regression forest (MT-RF). The conventional single-target regression forest (ST-RF) is also evaluated for comparison.

1) Parameter Setting: Table I shows the parameters for training our proposed image synthesis model for both the CT-to-MRI mapping and the MRI-to-CT mapping. The two directional synthesis tasks share mostly the same parameters. For each image synthesis direction, the Target 1 and Target 2 denote the output and input modalities, respectively. Specifically, the weights of the two regression targets are equally set to 0.5, as the anatomical details of the two modalities are complementary to each other. The size of the input image patch (used for feature extraction) is larger for the CT-to-MRI mapping, in that the task is a "simple-to-complex" mapping

TABLE I

PARAMETERS OF MULTI-TARGET REGRESSION FOREST (MT-RF) FOR CT-TO-MRI MAPPING AND MRI-TO-CT MAPPING

Parameter	CT-to-MRI mapping	MRI-to-CT mapping
λ	0.5, 0.5	
Input patch size	15×15×15	11×11×11
Feature #	750 (Haar-like features)×3 (image levels)= 2250	
Target patch size	Target1: 3×3×3; Target2: 3×3×3	
Tree number	24	20
Tree Depth	23	19
Auto-context	2 layers	

TABLE II

MEAN MAE AND PSNR OF 20 SUBJECTS WITH STANDARD DEVIATION FOR MRI-TO-CT MAPPING

	ST-RF	MT-RF	Auto-context model	
			2-layer MT-RF	3-layer MT-RF
MAE	51.64±3.83	45.69±3.52	39.47±3.41	**37.04±3.24**
PSNR (dB)	31.68±0.91	32.04±0.89	33.62±0.84	**34.03±0.82**

TABLE III

MEAN MAE AND PSNR OF 20 SUBJECTS WITH STANDARD DEVIATION FOR CT-TO-MRI MAPPING

	ST-RF	MT-RF	Auto-context model	
			2-layer MT-RF	3-layer MT-RF
MAE	140.37±10.32	136.72±9.63	124.11±8.70	**122.03±7.89**
PSNR(dB)	25.44±0.89	25.87±0.89	26.31±0.82	**26.52±0.78**

and generally more difficult than the "complex-to-simple" MRI-to-CT mapping. Considering more anatomical details are often available in MRI, we particularly adopt more and deeper trees in order to guarantee the robustness and accuracy of the CT-to-MRI synthesis. Here, we perform leave-one-out cross-validation and use two-layer auto-context model to refine the synthesis performance.

2) Performance Evaluation: Mean Absolute Error (MAE) and Peak Signal-to-Noise Ratio (PSNR) are utilized to evaluate the performance of image synthesis for both CT-to-MRI mapping and MRI-to-CT mapping. A synthesized image with higher quality should have lower MAE and higher PSNR values. In order to get the comparable results for both MRI-to-CT mapping and CT-to-MRI mapping, we normalize the intensity range of MRI to 0∼4000, which is the same to CT.

3) Image Synthesis Results: The quantitative results for both MRI-to-CT mapping and CT-to-MRI mapping are shown in Table II and Table III, respectively. For both MAE and PSNR values, the proposed **MT-RF** method outperforms **ST-RF**. The results demonstrate that our proposed MT-RF method can enhance the learning capability to build more accurate and robust mapping between CT and MRI. Moreover, the image synthesis performance can be further improved by using the auto-context model of cascading multiple layers of MT-RFs. The results are consistently improved for both MRI-to-CT mapping and CT-to-MRI mapping. The better measures, as well as the reduced standard deviation, demonstrate that using more neighborhood information as context

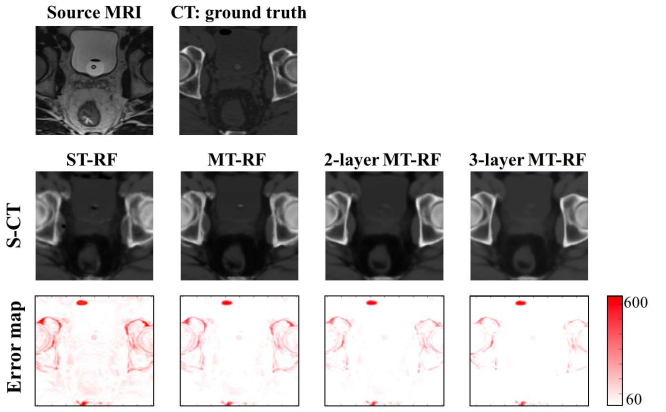


Fig. 8. Demonstration of synthesized CT (S-CT) from MRI. *First row*: source MRI and the ground-truth CT. *Second row*: the synthesized CT by using different methods evaluated in Table II. *Third row*: the residual maps of the absolute errors of the synthesized results.

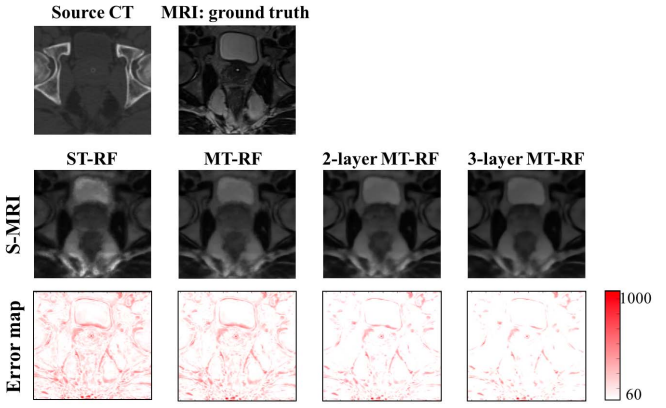


Fig. 9. Demonstration of synthesized MRI (S-MRI) from CT. *First row*: source CT and the ground-truth MRI. *Second row*: the synthesized MRI by using different methods evaluated in Table III. *Third row*: the residual maps of the absolute errors of the synthesized results.

features can help improve the synthesis performance. Note that the improvement on including the second-layer MT-RF is more significant than including the third layer, which means fast convergence of the auto-context model.

The visualized results of the synthesized CT and MRI are shown in Fig. 8 and Fig. 9, respectively. We can observe that the quality of the synthesized image is improved by using MT-RF with auto-context model. Although the synthesized image is smoother than the ground-truth image, however, all the anatomical details are clear and preserved well. It is reasonable that the MRI-to-CT mapping has higher performance than the CT-to-MRI mapping, as the appearance information in MRI is more complex than CT. However, the CT-to-MRI mapping (although more challenging) can still provide enough anatomical information for the subsequent registration, as the synthesized MRI is consistent with the actual MRI, especially around the organ boundaries.

TABLE IV

COMPARISON OF IMAGE REGISTRATION RESULTS IN TERMS OF DICE (%) VIA DIFFERENT IMAGE SYNTHESIS MANNERS BY USING THE PROPOSED REGISTRATION METHOD. A) WITHOUT SYNTHESIS: DIRECTLY REGISTERING THE ORIGINAL CT AND MRI (CT & MRI); B) SINGLE-DIRECTIONAL IMAGE SYNTHESIS: REGISTERING THE SAME MODALITY PAIR (CT & S-CT OR MRI & S-MRI); C) PROPOSED BI-DIRECTIONAL IMAGE SYNTHESIS (PROPOSED): USING TWO MODALITY PAIRS TO PERFORM REGION-ADAPTIVE REGISTRATION. "$*$" INDICATES STATISTICALLY SIGNI-FICANT IMPROVEMENT BY OUR PROPOSED REGISTRATION METHOD BASED ON BI-DIRECTIONAL IMAGE SYNTHESIS. ($p < 0.05$)

DICE (%)		Without synthesis	Single-directional		Bi-directional
		CT&MRI	CT&S-CT	MRI&S-MRI	Proposed
Intra-	Pro	86.4±5.7*	87.3±6.0*	88.1±4.2*	**89.5±2.3**
	Bla	89.7±2.5*	90.5±1.3*	91.9±0.7*	**93.4±0.4**
	Rec	79.1±4.0*	82.3±3.6*	85.3±2.8*	**87.6±2.4**
Inter-	Pro	66.7±8.9*	70.5±8.1*	72.4±5.9*	**76.3±3.4**
	Bla	78.0±9.5*	82.5±5.2*	82.9±5.0*	**84.7±4.6**
	Rec	63.6±14.1*	69.9±13.2*	70.8±11.9*	**73.5±7.5**

Pro: prostate; Bla: bladder; Rec: rectum. Intra-: Intra-subject; Inter-: Inter-subject

C. Experiments on Registration

In this section, the experiments are performed on both intra- and inter-subject CT/MRI registration. For *intra-subject* registration, we perform leave-one-out cross-validation. Every time the image synthesis model is trained from the 19 selected subjects. The registration is then validated upon the remaining one subject. For *inter-subject* registration, we first randomly select one patient and treat his CT image as the template. Then, we perform leave-one-out cross-validation upon the remaining 19 patients, by training the image synthesis model from 18 patients and validating the registration of the remaining one patient (using MRI) with the selected template (using CT). The above procedure is repeated by selecting 5 different template patients and the averaged results are reported.

1) Performance Evaluation: Based on the manual labels (L_M) of the three main pelvic organs (i.e., prostate, bladder and rectum) and the warped labels (L_W) after registration, two categories of measurement are utilized to evaluate the registration performance. **a) Volumetric overlap:** Dice Similarity Coefficient (DICE), Sensitivity (SEN) and Positive Predictive Value (PPV). **b) Surface distance:** Symmetric Average Surface Distance (SASD) and Hausdorff Distance (HAUS). The definitions of these measurements can be found in [45]. An accurate registration result should have higher volumetric overlap while lower surface distance.

2) Contribution of Learning-Based Image Synthesis: To evaluate the contribution of the image synthesis for multi-modal image registration, we provide Table IV to compare the registration results based on different image synthesis manners: a) **without image synthesis**, as we directly perform the registration on the original CT and MR images; b) **single-directional image synthesis**, as we perform registration on *either* CT modality pair (CT and S-CT) *or* MRI modality pair (MRI and S-MRI); and c) the **bi-directional image synthesis**, as we use both modality pairs and the proposed region-adaptive manner in this paper. Compared with the case of directly registering the original CT and MRI without image synthesis, the registration based on single-directional image synthesis improves the results as shown in Table IV. Moreover, the best performance is achieved by our proposed method based on the bi-directional image synthesis. That means, by using the complementary anatomical information from both modalities, registration accuracy and robustness can be further improved.

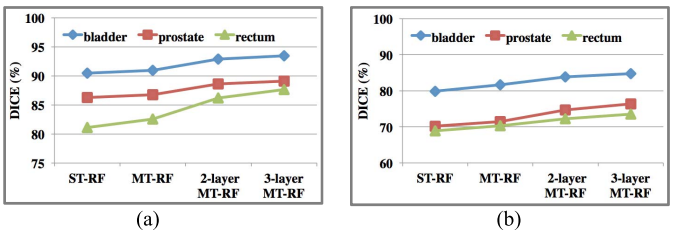


Fig. 10. MRI and CT pelvic image registration results (measured via Dice) by using our proposed registration method based on different image synthesis manners: 1) the single-target regression forest (**ST-RF**) and 2) our proposed multi-target regression forest (**MT-RF**) with refinement by multi-layer auto-context model (i.e., **2-layer MT-RF** and**3-layer MT-RF**), for (a) intra-subject registration and (b) inter-subject registration, respectively.

Fig. 10 presents the registration results of our proposed method by using ST-RF and MT-RF for image synthesis, respectively. From the results we can observe that, compared to ST-RF, our proposed MT-RF method can generate higher-quality synthesized images and thus boost the registration performance. The performance is further improved by utilizing the auto-context model (i.e., with multi-layer refinement in the 2-layer MT-RF and 3-layer MT-RF), as the quality of the synthesized images becomes even better. In general, the better quality of the synthesized CT and MRI (i.e., with clearer anatomy and less noise) is, the more accurate registration we can get in the final.

3) Contribution of Region-Adaptive Registration Strategy: In this subsection, we evaluate the contribution of our proposed region-adaptive strategy, i.e., selecting key points adaptively form CT and MRI, in helping steer the multi-modal pelvic image registration. To do this, we compare **our full method** with a) its variant without using the region-adaptive strategy and b) our previous multi-modal image registration method [38], both of which are described below.

a) We call our method without the region-adaptive strategy as **EQ** in this paper. Specifically, CT and MRI are equally treated to guide the registration, which is similar to the conventional multi-channel registration methods [57], [58]. Key points are selected based on the importance information from both CT and MRI. Then, for each selected key point, its local patches in both CT and MRI are extracted and stacked together as the morphological signature to guide the correspondence detection and deformable registration. Note that all other strategies in our full method are adopted for EQ.

b) Our previous multi-modal image registration [38] registers the CT pair and the MRI pair separately, and then perform an iterative dual-core deformation fusion (DDF) to combine the two deformation fields. Note that, for each modality pair, we use the same hierarchical symmetric registration as used in our full method, which is yet different from [38]. This implementation is called **DDF** in this paper.

TABLE V

COMPARISON OF MULTI-MODAL PELVIC IMAGE REGISTRATION RESULTS IN TERM OF DICE (%) BY THREE REGISTRATION METHODS, I.E., USING REGION-ADAPTIVE STRATEGY (OUR FULL METHOD) AND WITHOUT USING REGION-ADAPTIVE STRATEGY (EQ AND DDF). "∗" INDICATES STATISTICALLY SIGNIFICANT IMPROVEMENT BY OUR FULL METHOD ($p < 0.05$)

DICE (%)		EQ	DDF	Our full method
Intra-	Pro	88.3±3.9*	89.0±2.5*	**89.5±2.3**
	Bla	91.9±1.2*	92.8±0.8*	**93.4±0.4**
	Rec	86.1±4.6*	**88.1±2.6**	87.6±2.4
Inter-	Pro	69.8±5.8*	74.9±3.6*	**76.3±3.4**
	Bla	82.9±7.6*	82.2±5.2*	**84.7±4.6**
	Rec	70.8±9.3*	72.3±7.9*	**73.5±7.5**

TABLE VI

COMPARISON OF THE DICE (%) VALUES AFTER DEFORMABLE REGISTRATION BY TWO STATE-OF-THE-ART METHODS (SYN AND ITER-DDF) AND OUR FULL METHOD. "∗" INDICATES STATISTICALLY SIGNIFICANT IMPROVEMENT BY OUR FULL METHOD ($p < 0.05$)

DICE (%)		SyN [49] (MI)	Iter-DDF[38]		Our full method
			Demons	SyN	
Intra-	Pro	86.8±3.5*	88.9±4.3*	89.2±2.8*	**89.5±2.3**
	Bla	90.4±0.4*	93.2±0.5*	93.0±0.3*	**93.4**±0.4
	Rec	83.6±4.7*	86.6±2.5*	87.2±3.2*	**87.6±2.4**
Inter-	Pro	67.2±7.8*	73.9±4.1*	73.7±3.9*	**76.3±3.4**
	Bla	78.4±8.4*	81.9±5.7*	81.1±5.2*	**84.7±4.6**
	Rec	65.1±12.5*	69.2±8.4*	70.8±8.9*	**73.5±7.5**

Both of EQ and DDF are implemented upon the same bi-directional image synthesis, which are the same with our full method for fair comparison.

Table V compares registration results by the three methods. For *inter-subject* registration, the performances are significantly improved by **our full method** for all three pelvic organs. This shows that, by using the region-adaptive strategy, all the distinctive anatomical details in each modality can be retained for accurate correspondence detection, thus boosting the registration accuracy. In **EQ**, the way of equally treating the patches from both modalities can affect the anatomical distinctiveness, thus decreasing registration accuracy especially for the inter-subject case (with high image appearance/shape variation). From *intra-subject* registration, **our full method** can still beat the competing methods for both prostate and bladder as shown in Table V. For rectum, **DDF** works slightly better, but DDF is less efficient than our full method, since it needs to perform registration for multiple times. Overall, the proposed region-adaptive strategy is effective in enhancing deformable registration accuracy, especially for the case with large local deformations.

4) Comparison With Conventional Multi-Modal Deformable Registration Methods: Two state-of-the-art multi-modal deformable registration methods are selected for comparison. The first one is Symmetric Image Normalization (**SyN**) [49], which can perform symmetric registration by using MI as the similarity metric [59]. This method is ranked best among 14 state-of-the-art registration methods for brain MR image registration in [60]. The second one is our previous method [38], namely **Iter-DDF**, using iterative dual-core deformation fusion. Iter-DDF may use Demons [54] or SyN for registration after image synthesis; thus the results are separately reported.

TABLE VII

COMPARISON OF THE MEAN SEN, PPV, SASD AND HAUS VALUES OF THREE PELVIC ORGANS AFTER DEFORMABLE REGISTRATION BY THE TWO STATE-OF-THE-ART METHODS (SYN AND ITER-DDF), AND OUR FULL METHOD, RESPECTIVELY. "∗" INDICATES STATISTICALLY SIGNIFICANT IMPROVEMENT BY OUR FULL METHOD ($p < 0.05$)

Metric		SyN [49] (MI)	Iter-DDF[38]		Our full method
			Demons	SyN	
SEN (%)	Intra-	81.1±6.5*	86.6±6.3*	86.4±5.2	**86.8±3.7**
	Inter-	67.1±7.9*	79.9±8.2*	79.7±7.1*	**84.8±6.3**
PPV (%)	Intra-	91.7±2.2*	95.2±1.8*	95.4±1.7*	**95.6±1.2**
	Inter-	72.3±10.8*	72.9±5.1*	73.1±4.9*	**76.2±4.5**
SASD (mm)	Intra-	1.21±0.94*	1.03±0.64	1.10±0.71*	**1.02±0.57**
	Inter-	2.45±1.84*	1.71±0.74*	1.67±0.76*	**1.59±0.70**
HAUS (mm)	Intra-	9.06±4.25*	6.71±2.30*	6.69±1.94*	**5.42±1.68**
	Inter-	13.88±5.70*	10.56±5.06*	10.72±4.65*	**10.63±4.40**

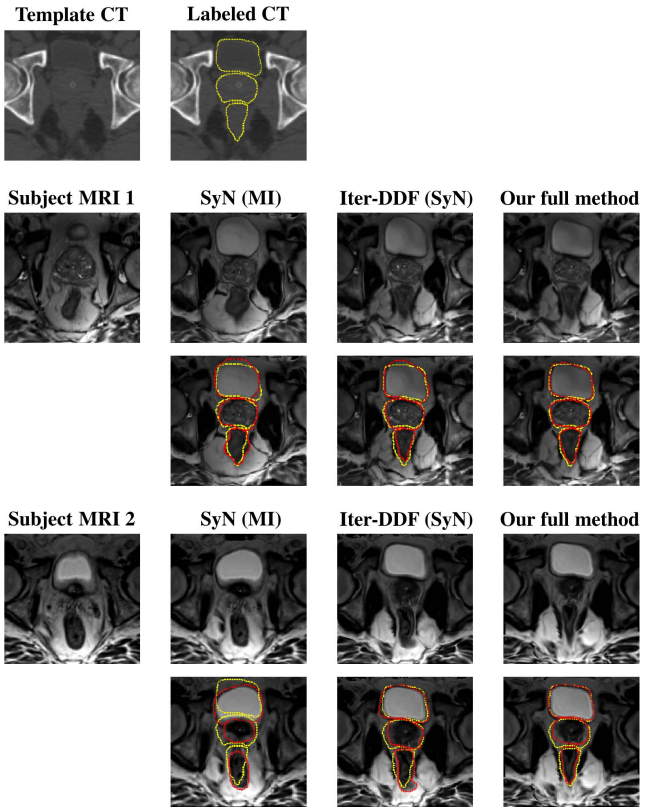


Fig. 11. Demonstration of inter-subject registration results by two state-of-the-art methods (SyN (MI) and Iter-DDF (SyN)) and our full method. Here, yellow contours denote the manually-labeled boundaries of bladder, prostate and rectum in the template CT, red contours denote the respective manually-labeled organ boundaries in MRI after registration.

Table VI shows the registration results obtained by SyN, Iter-DDF and our full method. The direct registration of CT and MRI using MI in SyN can achieve reasonable results, as shown in Table VI. However, its accuracy is limited, especially for the highly challenging inter-subject registration case. Iter-DDF obtains better registration performance than SyN by using bi-directional image synthesis. While, our full method achieves the best performance, especially for inter-subject registration. This demonstrates that our full method can better tackle the challenges of large local deforma-

tions, which is crucial for accurate inter-subject pelvic image registration.

We further provide other overlap ratios (SEN and PPV), and surface distances (SASD and HAUS) in Table VII, by averaging over three organs. These results show that **our full method** can consistently produce higher SEN and PPV while lower SASD and HAUS, compared with state-of-the-art methods. For the *inter-subject* case, the improvement is more significant than the *intra-subject* case. It is difficult for conventional registration method to accurately handle the regions with large shape variation, e.g., bladder and rectum. With the region-adaptive and the hierarchical symmetric strategies in our method, these limitations can be effectively addressed.

Fig. 11 provides typical *inter-subject* registration results for visual comparison, by **SyN (MI)**, **Iter-DDF (SyN)**, and **our full method**. As can be seen from the first column of Fig. 11, the original deformations of bladder, prostate and rectum across subjects are quite large. Thus, it is quite difficult to directly register the original CT and MRI via **SyN**, as shown in the second column. **Iter-DDF (SyN)** can provide much better results, as shown in the third column. Obviously, **our full method** achieves the best registration performance, indicated by the largest organ overlap with the labels of the template CT in the last column. Besides, the morphologies of all pelvic organs are well preserved, i.e., with no obvious distortions in the warped MRI, which indicates the superiority of our full method.

IV. Discussion and Conclusion

In order to use the complementary anatomical information from both modalities for guiding accurate multi-modal image registration, we propose to use bi-directional image synthesis, i.e., synthesizing a) CT from MRI and b) also MRI from CT. For CT-to-MRI synthesis, it is difficult as it maps the CT modality with limited anatomical information to the MRI with rich anatomical information. To this end, we propose MT-RF to solve the bi-directional image synthesis, which well preserve anatomical details in image synthesis to boost the performance of subsequent registration.

Even for the pelvic images of the same patient, the shapes of bladder and rectum change dynamically across time, which may subsequently affect the location/shape of the prostate. Thus, the local deformations in pelvic images are highly challenging for image registration. In this paper, we have proposed a region-adaptive registration method in a hierarchical and symmetric fashion. The benefits of this region-adaptive registration method have three folds. **a)** The key points own discriminative anatomical information and guide more robust registration than incorporating all voxels and their intensities. **b)** The region-adaptive strategy allows detecting anatomical correspondences by adaptively fusing the two modalities. **c)** The use of the hierarchical and symmetric registration further improves the registration accuracy, especially for the case of inter-subject registration with large deformations.

In this paper, we have proposed a new region-adaptive registration method, based on learning-based image synthesis, for pelvic CT and MRI registration, which is crucial for radiation therapy of prostate cancer. A novel MT-RF is introduced to first bridge the appearance gap between pelvic CT and MRI in bi-directions. Then, to make full use of the synthesized CT and MRI, we further propose to extract key points from both CT and MRI adaptively and steer the deformable registration in a hierarchical and symmetric fashion. Experimental results show the contribution of each proposed strategy, and also the superiority of our proposed region-adaptive registration method over the state-of-the-art multi-modal deformable registration methods, using various performance metrics.

References

[1] R. de Crevoisier *et al.*, "Increased risk of biochemical and local failure in patients with distended rectum on the planning CT for prostate cancer radiotherapy," *Int. J. Radiat. Oncol. Biol. Phys.*, vol. 62, no. 4, pp. 965–973, 2005.

[2] P. Metcalfe, "The potential for an enhanced role for MRI in radiation-therapy treatment planning," *Technol. Cancer Res. Treatment*, vol. 12, no. 5, pp. 429–446, 2013.

[3] 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.

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

[5] 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.

[6] C. Guetter, C. Xu, F. Sauer, and J. Hornegger, "Learning based non-rigid multi-modal image registration using Kullback–Leibler divergence," in *Proc. Int. Conf. Med. Image Comput. Comput.-Assist. Intervent.*, 2005, pp. 255–262.

[7] R. W. K. So and A. C. S. Chung, "Learning-based multi-modal rigid image registration by using Bhattacharyya distances," in *Proc. Annu. Int. Conf. IEEE Eng. Med. Biol. Soc.*, Aug./Sep. 2011, pp. 2642–2645.

[8] R. Liao, C. Guetter, C. Xu, Y. Sun, A. Khamene, and F. Sauer, "Learning-based 2D/3D rigid registration using Jensen–Shannon divergence for image-guided surgery," in *Proc. Int. Workshop Med. Imag. Virtual Reality*, 2006, pp. 228–235.

[9] 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.

[10] A. Andronache, M. von Siebenthal, G. Székely, and P. Cattin, "Non-rigid registration of multi-modal images using both mutual information and cross-correlation," *Med. Image Anal.*, vol. 12, no. 1, pp. 3–15, 2008.

[11] J. Chappelow *et al.*, "Elastic registration of multimodal prostate MRI and histology via multiattribute combined mutual information," *Med. Phys.*, vol. 38, no. 4, pp. 2005–2018, 2011.

[12] Y. S. Kim, J. H. Lee, and J. B. Ra, "Multi-sensor image registration based on intensity and edge orientation information," *Pattern Recognit.*, vol. 41, no. 11, pp. 3356–3365, 2008.

[13] Z. F. Knops, J. B. A. Maintz, M. A. Viergever, and J. P. W. Plui, "Normalized mutual information based registration using k-means clustering and shading correction," *Med. Image Anal.*, vol. 10, no. 3, pp. 432–439, 2006.

[14] J. Liang, X. Liu, K. Huang, X. Li, D. Wang, and X. Wang, "Automatic registration of multisensor images using an integrated spatial and mutual information (SMI) metric," *IEEE Trans. Geosci. Remote Sens.*, vol. 52, no. 1, pp. 603–615, Jan. 2014.

[15] H. Rivaz, Z. Karimaghaloo, V. S. Fonov, and D. L. Collins, "Nonrigid registration of ultrasound and MRI using contextual conditioned mutual information," *IEEE Trans. Med. Imag.*, vol. 33, no. 3, pp. 708–725, Mar. 2014.

[16] J. P. W. Pluim, J. B. A. Maintz, and M. A. Viergever, "Image registration by maximization of combined mutual information and gradient information," in *Proc. Int. Conf. Med. Image Comput. Comput.-Assist. Intervention*, 2000, pp. 452–461.

[17] D. Loeckx, P. Slagmolen, F. Maes, D. Vandermeulen, and P. Suetens, "Nonrigid image registration using conditional mutual information," *IEEE Trans. Med. Imag.*, vol. 29, no. 1, pp. 19–29, Jan. 2010.

[18] X. Zhuang, S. Arridge, D. J. Hawkes, and S. Ourselin, "A nonrigid registration framework using spatially encoded mutual information and free-form deformations," *IEEE Trans. Med. Imag.*, vol. 30, no. 10, pp. 1819–1828, Oct. 2011.

[19] J. Woo, M. Stone, and J. L. Prince, "Multimodal registration via mutual information incorporating geometric and spatial context," *IEEE Trans. Image Process.*, vol. 24, no. 2, pp. 757–769, Feb. 2015.

[20] P. A. Legg, P. L. Rosin, D. Marshall, and J. E. Morgan, "Feature neighbourhood mutual information for multi-modal image registration: An application to eye fundus imaging," *Pattern Recognit.*, vol. 48, no. 6, pp. 1937–1946, 2015.

[21] Y. Ou, A. Sotiras, N. Paragios, and C. Davatzikos, "DRAMMS: Deformable registration via attribute matching and mutual-saliency weighting," *Med. Image Anal.*, vol. 15, no. 4, pp. 622–639, 2011.

[22] T. Brox and J. Malik, "Large displacement optical flow: Descriptor matching in variational motion estimation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 33, no. 3, pp. 500–513, Mar. 2011.

[23] D. Shen and C. Davatzikos, "HAMMER: Hierarchical attribute matching mechanism for elastic registration," *IEEE Trans. Med. Imag.*, vol. 21, no. 11, pp. 1421–1439, Nov. 2002.

[24] G. Wu, F. Qi, and D. Shen, "Learning best features for deformable registration of MR brains," in *Proc. Int. Conf. Med. Image Comput. Comput.-Assisted Intervent.*, 2005, pp. 179–187.

[25] G. Wu, F. Qi, and D. Shen, "Learning-based deformable registration of MR brain images," *IEEE Trans. Med. Imag.*, vol. 25, no. 9, pp. 1145–1157, Sep. 2006.

[26] D. G. Lowe, "Object recognition from local scale-invariant features," in *Proc. 7th IEEE Int. Conf. Comput. Vis.*, Sep. 1999, pp. 1150–1157.

[27] K. Mikolajczyk and C. Schmid, "A performance evaluation of local descriptors," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 27, no. 10, pp. 1615–1630, Oct. 2005.

[28] M. P. Heinrich *et al.*, "MIND: Modality independent neighbourhood descriptor for multi-modal deformable registration," *Med. Image Anal.*, vol. 16, no. 7, pp. 1423–1435, 2012.

[29] A. Johansson, M. Karlsson, and T. Nyholm, "CT substitute derived from MRI sequences with ultrashort echo time," *Med. Phys.*, vol. 38, no. 5, pp. 2708–2714, 2011.

[30] S. Roy, W.-T. Wang, A. Carass, J. L. Prince, J. A. Butman, and D. L. Pham, "PET attenuation correction using synthetic CT from ultrashort echo-time MR imaging," *J. Nucl. Med.*, vol. 55, no. 12, pp. 2071–2077, 2014.

[31] A. Jog, A. Carass, and J. L. Prince, "Improving magnetic resonance resolution with supervised learning," in *Proc. IEEE 11th Int. Symp. Biomed. Imag. (ISBI)*, Apr./May 2014, pp. 987–990.

[32] T. Huynh *et al.*, "Estimating CT image from MRI data using structured random forest and auto-context model," *IEEE Trans. Med. Imag.*, vol. 35, no. 1, pp. 174–183, Jan. 2016.

[33] R. Li *et al.*, "Deep learning based imaging data completion for improved brain disease diagnosis," in *Proc. Int. Conf. Med. Image Comput. Comput.-Assist. Intervention*, 2014, pp. 305–312.

[34] D. Nie, X. Cao, Y. Gao, L. Wang, and D. Shen, "Estimating CT image from MRI data using 3D fully convolutional networks," in *Proc. Int. Workshop Large-Scale Annotation Biomed. Data Expert Label Synth.*, 2016, pp. 170–178.

[35] A. Roche, X. Pennec, G. Malandain, and N. Ayache, "Rigid registration of 3-D ultrasound with MR images: A new approach combining intensity and gradient information," *IEEE Trans. Med. Imag.*, vol. 20, no. 10, pp. 1038–1049, Oct. 2001.

[36] W. Wein, S. Brunke, A. Khamene, M. R. Callstrom, and N. Navab, "Automatic CT-ultrasound registration for diagnostic imaging and image-guided intervention," *Med. Image Anal.*, vol. 12, no. 5, pp. 577–585, 2008.

[37] S. Roy, A. Carass, A. Jog, J. L. Prince, and J. Lee, "MR to CT registration of brains using image synthesis," *Proc. SPIE, Med. Imag., Image Process.*, vol. 9034, p. 903419, Mar. 2014.

[38] X. Cao, Y. Gao, J. Yangm, G. Wu, and D. Shen, "Learning-based multimodal image registration for prostate cancer radiation therapy," in *Proc. Int. Conf. Med. Image Comput. Comput.-Assist. Intervent.*, 2016, pp. 1–9.

[39] X. Cao, J. Yang, Y. Gao, Y. Guo, G. Wu, and D. Shen, "Dual-core steered non-rigid registration for multi-modal images via bi-directional image synthesis," *Med. Image Anal.*, vol. 41, pp. 18–31, Oct. 2017.

[40] A. Liaw and M. Wiener, "Classification and regression by randomforest," *R News*, vol. 2, no. 3, pp. 18–22, 2002.

[41] J. Zhang, Y. Gao, S. H. Park, X. Zong, W. Lin, and D. Shen, "Structured learning for 3-D perivascular space segmentation using vascular features," *IEEE Trans. Biomed. Eng.*, vol. 64, no. 12, pp. 2803–2812, Dec. 2017.

[42] J. Zhang, M. Liu, and D. Shen, "Detecting anatomical landmarks from limited medical imaging data using two-stage task-oriented deep neural networks," *IEEE Trans. Image Process.*, vol. 26, no. 10, pp. 4753–4764, Oct. 2017.

[43] L. Wei *et al.*, "Learning-based deformable registration for infant MRI by integrating random forest with auto-context model," *Med. Phys.*, vol. 44, no. 12, pp. 6289–6303, 2017.

[44] A. Criminisi and J. Shotton, *Decision Forests for Computer Vision and Medical Image Analysis*. Springer, 2013.

[45] Y. Gao, Y. Shao, J. Lian, A. Z. Wang, R. C. Chen, and D. Shen, "Accurate segmentation of CT male pelvic organs via regression-based deformable models and multi-task random forests," *IEEE Trans. Med. Imag.*, vol. 35, no. 6, pp. 1532–1543, Jun. 2016.

[46] P. Viola and M. J. Jones, "Robust real-time face detection," *Int. J. Comput. Vis.*, vol. 57, no. 2, pp. 137–154, 2004.

[47] Z. Tu and X. Bai, "Auto-context and its application to high-level vision tasks and 3D brain image segmentation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 32, no. 10, pp. 1744–1757, Oct. 2010.

[48] X. Cao *et al.*, "Deformable image registration based on similarity-steered CNN regression," in *Proc. Int. Conf. Med. Image Comput. Comput.-Assist. Intervention*. 2017, pp. 300–308.

[49] B. B. Avants, C. L. Epstein, M. Grossman, and J. C. Gee, "Symmetric diffeomorphic image registration with cross-correlation: Evaluating automated labeling of elderly and neurodegenerative brain," *Med. Image Anal.*, vol. 12, no. 1, pp. 26–41, 2008.

[50] B. B. Avants, M. Grossman, and J. C. Gee, "Symmetric diffeomorphic image registration: Evaluating automated labeling of elderly and neurodegenerative cortex and frontal lobe," in *Proc. Int. Workshop Biomed. Image Registration*, 2006, pp. 50–57.

[51] G. Wu, M. Kim, Q. Wang, and D. Shen, "S-HAMMER: Hierarchical attribute-guided, symmetric diffeomorphic registration for MR brain images," *Human Brain Mapping*, vol. 35, no. 3, pp. 1044–1060, 2014.

[52] G. Wu, P.-T. Yap, M. Kim, and D. Shen, "TPS-HAMMER: Improving HAMMER registration algorithm by soft correspondence matching and thin-plate splines based deformation interpolation," *NeuroImage*, vol. 49, no. 3, pp. 2225–2233, 2010.

[53] F. L. Bookstein, "Principal warps: Thin-plate splines and the decomposition of deformations," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 11, no. 6, pp. 567–585, Jun. 1989.

[54] T. Vercauteren, X. Pennec, A. Perchant, and N. Ayache, "Diffeomorphic demons: Efficient non-parametric image registration," *NeuroImage*, vol. 45, no. 1, pp. S61–S72, 2009.

[55] J. G. Sled, A. P. Zijdenbos, and A. C. Evans, "A nonparametric method for automatic correction of intensity nonuniformity in MRI data," *IEEE Trans. Med. Imag.*, vol. 17, no. 1, pp. 87–97, Feb. 1998.

[56] M. Jenkinson and S. Smith, "A global optimisation method for robust affine registration of brain images," *Med. Image Anal.*, vol. 5, no. 2, pp. 143–156, Jun. 2001.

[57] Y. Li and R. Verma, "Multichannel image registration by feature-based information fusion," *IEEE Trans. Med. Imag.*, vol. 30, no. 3, pp. 707–720, Apr. 2011.

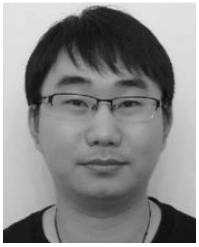
[58] D. Forsberg, Y. Rathi, S. Bouix, D. Wassermann, H. Knutsson, and C.-F. Westin, "Improving registration using multi-channel diffeomorphic demons combined with certainty maps," in *Proc. Int. Workshop Multimodal Brain Image Anal.*, 2011, pp. 19–26.

[59] 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, 2011.

[60] A. Klein *et al.*, "Evaluation of 14 nonlinear deformation algorithms applied to human brain MRI registration," *NeuroImage*, vol. 46, no. 3, pp. 786–802, 2009.

Xiaohuan Cao received the bachelor's and master's degrees from Northwestern Polytechnical University, where she is currently pursuing the Ph.D. degree with the School of Automation. She is also a visiting student in The University of North Carolina at Chapel Hill, Chapel Hill, NC, USA. Her research interests include machine learning and medical image analysis.

Jianhua Yang received the bachelor's degree from Xidian University and the master's and Ph.D. degrees from Northwestern Polytechnical University. She is currently a Professor with the School of Automation, Northwestern Polytechnical University. Her research interests include biomedical image processing, detection, and control technology.



Yaozong Gao received the Ph.D. degree from the Department of Computer Science, The University of North Carolina at Chapel Hill. He was a Computer Vision Researcher with Apple. He is currently directing the Deep Learning Group, United Imaging Intelligence, China. He has published over 90 papers in the international journals and conferences, such as MICCAI, TMI, and MIA. His research interests include machine learning, computer vision, and medical image analysis.

Qian Wang received the Ph.D. degree in computer science from The University of North Carolina at Chapel Hill in 2013. He is currently the Director of the Medical Image Computing Laboratory, Institute of Medical Imaging Technology, School of Biomedical Engineering, Shanghai Jiao Tong University. His researches interests include medical image analysis, computer vision, machine learning, artificial intelligence, and translational medical studies.

Dinggang Shen (F'18) is currently a Jeffrey Houpt Distinguished Investigator, and a Professor of Radiology, Computer Science, and Biomedical Engineering, with the Biomedical Research Imaging Center (BRIC), The University of North Carolina at Chapel Hill. He is currently directing the Image Display, Enhancement, and Analysis Lab, Center for Image Analysis and Informatics, Department of Radiology, and also the medical image analysis core at the BRIC. He was a tenure-track Assistant Professor with the University of Pennsylvanian, Philadelphia, PA, USA, and a Faculty Member with the Johns Hopkins University, Baltimore, MD, USA. He has authored over 800 papers in the international journals and conference proceedings. His research interests include medical image analysis, computer vision, and pattern recognition. He is a fellow of The American Institute for Medical and Biological Engineering, and also a fellow of The International Association for Pattern Recognition. He serves as an Editorial Board Member for eight international journals. He has also served on the Board of Directors, The Medical Image Computing and Computer Assisted Intervention Society, from 2012 to 2015.