



Published in final edited form as:

Med Phys. 2025 February ; 52(2): 938–950. doi:10.1002/mp.17536.

Tumor aware recurrent inter-patient deformable image registration of computed tomography scans with lung cancer

Jue Jiang,

Chloe Min Seo Choi,

Maria Thor,

Joseph O. Deasy,

Harini Veeraraghavan*

Department of Medical Physics, Memorial Sloan Kettering Cancer Center, NY, New York, USA.

Abstract

Background: Voxel-based analysis (VBA) for population level radiotherapy (RT) outcomes modeling requires topology preserving inter-patient deformable image registration (DIR) that preserves tumors on moving images while avoiding unrealistic deformations due to tumors occurring on fixed images.

Purpose: We developed a TumoR-Aware reCurrent Registration (TRACER) deep learning (DL) method and evaluated its suitability for VBA.

Methods: TRACER consists of encoder layers implemented with stacked 3D convolutional long short term memory network (3D-CLSTM) followed by decoder and spatial transform layers to compute dense deformation vector field (DVF). Multiple CLSTM steps are used to compute a progressive sequence of deformations. Input conditioning was applied by including tumor segmentations with 3D image pairs as input channels. Bidirectional tumor rigidity, image similarity, and deformation smoothness losses were used to optimize the network in an unsupervised manner. TRACER and multiple DL methods were trained with 204 3D CT image pairs from patients with lung cancers (LC) and evaluated using (a) Dataset I (N = 308 pairs) with DL segmented LCs, (b) Dataset II (N = 765 pairs) with manually delineated LCs, and (c) Dataset III with 42 LC patients treated with RT.

Results: TRACER accurately aligned normal tissues. It best preserved tumors, indicated by the smallest tumor volume difference of 0.24%, 0.40%, and 0.13% and mean square error in CT intensities of 0.005, 0.005, 0.004, computed between original and resampled moving image tumors, for Datasets I, II, and III, respectively. It resulted in the smallest planned RT tumor dose difference computed between original and resampled moving images of 0.01 Gy and 0.013 Gy when using a female and a male reference.

Conclusions: TRACER is a suitable method for inter-patient registration involving lung cancers occurring in both fixed and moving images and applicable to voxel-based analysis methods.

* veerarah@mskcc.org .

Jue Jiang and Harini Veeraraghavan contributed equally.

I. Introduction

Technological advances in image guidance and radiotherapy (RT) has resulted in several new available treatment options. Voxel-based analysis (VBA) is a useful technique for studying population-level outcomes in radiation therapy (RT), such as radiation-induced toxicity. It involves comparing imaging data at the voxel level between different patients, and provides detailed quantification of how different regions within an organ respond to spillover radiation at a population level^{1,2}. However, VBA requires topology preserving deformable image registration (DIR) between different patient scans, or inter-patient DIR, so that voxels from different patients are brought to a common reference coordinate frame for comparing voxel-level dose sensitivities between patients. Accurate inter-patient DIR is thus essential for reliably accurate identification of dose-sensitive regions within organs using VBA.

VBA methods typically use iterative B-spline DIR^{1,2,3} to align static 3D scans of different patients to a reference anatomy, because these methods do not require disease specific training. Aligning different patient scans is challenging due to anatomy differences, wherein tumor present in one patient has no physiological or spatial correspondence with tumor occurring in another patient. If care is not taken to handle anatomically different image regions occurring in the moving (patient scan from studied population) and fixed image (reference scan), inaccurate registrations can occur, leading to dose response variability⁴, and exclusion of inaccurately aligned patient scans for VBA^{5,6,7}. VBA performed with less accurate inter-patient DIR method may not be representative of the broader population of patients with the studied disease, motivating our approach. We developed a novel deep learning (DL) DIR method to align static 3D computed tomography (CT) images of entirely different patients called Tumor-Aware recurrEnt Registration (TRACER). TRACER handles scenarios where lung tumors occur in non-corresponding spatial locations in the moving and fixed images. We focus on aligning static 3D instances of one patient anatomy to a reference patient anatomy.

Inter-patient DIR has been studied extensively in the context of atlas based registration^{8,9,10} for normal tissue segmentation. DL-DIR methods have typically focused on aligning patient scans without visible tumors^{11,12}. Although normal anatomy varies between patients, organs occupy fixed anatomic locations unlike tumors. Hence, presence of tumors adds additional complexity in achieving reliably accurate inter-patient registration, necessitating careful selection of the canonical patient for VBA. Groupwise registration methods can overcome the need for careful selection of canonical patient by aligning all patients in a group to an average reference^{13,14}. However, computational and memory requirements of this approach is prohibitive as all images in the group are required as inputs to the model.

Rigidity constraints are commonly used in DIR to preserve geometry of both normal tissues and tumors^{15,16} applied to intra-patient longitudinal registration, with an explicit assumption that no new tumors occur on fixed images. Most deep learning methods performing single step registration cannot preserve topology due to their assumption of a small deformation framework that is approximated using a stationary velocity field (SVF)¹⁷. Inverse consistency regularization using cyclical image registration¹⁸, inverse flow

generation constraints¹⁹, as well as bidirectional registration to an intermediate space¹⁷ are some example methods used to ensure topologically consistent deformations. However, such methods focus on achieving flow path invariant deformations, which doesn't ensure preservation of structures of interest. Similarly, recurrent networks^{20,21} can handle large anatomy differences by performing progressive deformation refinement but cannot preserve tumors and cause physically unrealistic tissue stretching in locations containing tumors on fixed images without tumor conditioning (Figure. 1(a)).

TRACER builds on our prior work called patient specific anatomic and shape context (PACS)²¹ for aligning intra-patient CT with during treatment cone beam CTs. Similar to PACS, it employs a 3D image registration network where the encoder layers consist of stacked 3D convolutional long short term memory network (3DCLSTM) used to progressively refine the extracted features and ultimately the deformations. TRACER differs from PACS by extending the use case to inter-patient DIR for image volumes with non-corresponding tumors. It introduces input tumor conditioning combined with bi-directional tumor rigidity implemented to preserve tumors in the forward flow and penalize distortions within tumor voxels occurring in fixed images in the inverse flow. Input tumor conditioning also relaxes rigidity penalty by allowing tissues adjacent to the conditioned structure to deform. Our approach can be fully automated by using auto-segmented tumors for conditioning the registration. TRACER thus addresses the shortcomings of prior methods and effectively preserves tumors while also avoiding unrealistic tissue stretching (Figure. 1 (b)). To our best knowledge, this is the first DL-DIR work to align different patient scans containing tumors in non-corresponding locations on moving and fixed images.

Our contributions are: (a) a new inter-patient recurrent DIR method that is robust to tumors occurring in both fixed and moving images, (b) tumor conditioned DIR to preserve geometry and CT intensity of tumors and prevent unrealistic distortions due to non-corresponding tumors occurring on fixed images, (c) fully automated approach using deep learning tumor segmentation to condition DIR, and (d) rigorous testing of our approach on three datasets in patients with locally advanced non-small cell lung cancer (LA-NSCLC) treated with RT.

II. Method

Problem setting and approach:

Given a pair of rigidly aligned static 3D CT scans from two different patients $\{I_m, I_f\}$, the goal is to deformably align moving I_m to fixed image I_f while preserving topology and all visible tumors occurring on moving image. This work is concerned with aligning two different and unrelated patient scans where tumors occurring in one patient or moving image I_m have no relation to the tumors occurring in a different patient or fixed image I_f . We achieve this goal by conditioning the registration using the tumor segmentation masks $\{y_m, y_f\}$ as additional input channels with $\{I_m, I_f\}$ to the model (Figure. 2). Input structure conditioning modulates the extracted image features, and tumor masks are also used when computing image similarity and bidirectional rigidity penalties.

II.A. Tumor-aware recurrent registration network (TRACER):

TRACER is a structure-conditioned registration method, which combines 3D input image pairs with corresponding 3D structure segmentation masks ($\{I_m, I_f, y_m, y_f\}$) as its inputs.

Our approach uses automated tumor segmentations produced using a published multiple resolution residual network (MRRN)²², available through open-source CERR software¹ (Figure. 2(a)). TRACER architecture (Figure. 2(b)) uses a 3D-Unet style architecture where each encoder layer consists of a 3DCLSTM layer followed by a convolutional layer with stride of 2 followed by LeakyReLU activation. In addition to progressively refining the extracted deformations, stacked 3D-CLSTM architecture allows continuity of computed features between the various feature levels.

The moving image (I_m), fixed image (I_f), and the segmentations (y_m and y_f) are input to the first CLSTM step at $t = 1$, with the hidden state initialized with zero. Subsequent CLSTM steps use the output consisting of the deformed moving image and the corresponding resampled segmentation and the fixed image and its segmentation as well as the hidden state updated from prior CLSTM step as their inputs (Figure. 2 (a)). Structure conditioning is explicitly implemented into the CLSTM layers by updating the gating filters consisting of forget gate f^t , memory cells c^t , hidden state h^t , input state i^t , and output gate o^t as below:

$$\begin{aligned}
 & \text{Input State for } x^t \\
 f^t &= \sigma(W_{xf} * \left\{ \begin{matrix} I_m^{t-1}, y_m^{t-1}, I_f, y_f \end{matrix} \right\} + W_{hf} * h^{t-1} + b_f) \\
 i^t &= \sigma(W_{xi} * \left\{ \begin{matrix} I_m^{t-1}, y_m^{t-1}, I_f, y_f \end{matrix} \right\} + W_{hi} * h^{t-1} + b_i) \\
 c^t &= \tanh(W_{xc} * \left\{ \begin{matrix} I_m^{t-1}, y_m^{t-1}, I_f, y_f \end{matrix} \right\} + W_{hc} * h^{t-1} + b_c) \\
 o^t &= \sigma(W_{xo} * \left\{ \begin{matrix} I_m^{t-1}, y_m^{t-1}, I_f, y_f \end{matrix} \right\} + W_{ho} * h^{t-1} + b_o) \\
 c^t &= f^t \odot c^{t-1} + i^t \odot c^t \\
 h^t &= o^t \odot \tanh(c^t),
 \end{aligned} \tag{1}$$

where, σ is the sigmoid activation function, $*$ the convolution operator, $\odot$ the Hadamard product, and W the weight matrix.

The decoder layers (same number as the encoder layers) of the registration network are composed of convolutional filters with up-sampling operations to extract the dense 3D flow field. Skip connections are implemented between the encoder and decoder layers to improve training stability and the sharpness of extracted deformations. The flow field is converted into deformation vector field (DVF) using spatial transform networks²³. Fast computation is achieved by assuming a stationary velocity field (SVF) v_j for the deformation computed at the end of each CLSTM step. Hence the DVF ϕ^j in each step j is computed

¹ <https://www.github.com/cerr/CERR>

as $\frac{d\phi_i^j}{dj} = v^j(\phi_i^j)$, subject to $\phi_{i=0}^j = Id$ using spatial time integration ($i = 7$)^{17,24,25} performed using scaling and squaring transforms⁸. Diffeomorphic integration layer is implemented to convert the incremental flow field into DVF while ensuring diffeomorphic deformation⁸. The velocity v_j is allowed to vary for each CLSTM step, which in turn allows computation of a temporally and spatially varying displacement field. Note that our computation using a SVF framework is performed for the incremental deformation computed for each CLSTM step as opposed to methods that use SVF for computing the entire deformation in a single step^{8,25}. Our approach thus ensures higher accuracy in the presence of large deformations. The progressive sequence of deformations to deform the moving image x_m to fixed image x_f in T CLSTM steps, with each CLSTM step using N time integrations is thus computed as, $\phi_m^f = \phi_{i=0}^T \circ \phi_{i=0}^N \phi_i^f(v^f)$, where $\phi_{i=0}^T \phi_i^f(x_m) = \phi^1 \circ (\phi^2 \dots \circ (\phi^T, x_m))$ is a composition operator.

II.B. Tumor-aware network regularization

The network is optimized by maximizing the posterior probability $P(\phi | I_f, y_f, I_m, y_m)$ of the forward dense deformation flow used to map the moving image (I_m) into the spatial coordinates of the fixed image (I_f). The posterior $P(\phi | I_f, y_f, I_m, y_m)$ can be decomposed as:

$$\log P(\phi | I_m, y_m, I_f, y_f) = \log P(I_f, y_f | I_m, y_m, \phi) + \log P(\phi | I_m, y_m) + \log P(y_m | I_m) - \log P(I_f, y_f | I_m, y_m) - \log P(y_m | I_m) \quad (2)$$

The terms (I_f, y_f) and (I_m, y_m) are independent and they come from two different patients that have no temporal or sequential relationship²⁶. Hence, Eqn. 2 can be simplified as:

$$\begin{aligned} &\approx \log P(I_f, y_f | I_m, y_m, \phi) + \log P(\phi | I_m, y_m) \\ &\approx \log P(I_f, y_f | I_m, y_m, \phi) + \log P(\phi | I_m) + \log P(\phi | y_m) + \log P(y_m | I_m). \end{aligned} \quad (3)$$

The term $P(y_m | I_m)$ is constant because the segmentation is computed prior to optimization using the DIR network. The posterior $P(\phi | I_f, y_f, I_m, y_m)$ can now be expressed as:

$$\begin{aligned} \log P(\phi | I_f, y_f, I_m, y_m) &= \underbrace{\log P(I_f, y_f | I_m, y_m, \phi)}_{\text{Similarity}} + \underbrace{\log P(\phi | I_m)}_{\text{Smoothness}} + \underbrace{\log P(\phi | y_m)}_{\text{Rigidity}} + \text{constant}. \\ &= L_{sim} + L_{smooth} + L_{tumor}^{pre} \end{aligned} \quad (4)$$

Image similarity (L_{sim}) in Eqn. 4 is computed by removing the effect of tumors in moving and fixed image as:

$$L_{sim} = \sum_{t=1}^N E \| I_f \odot (I - y_f), I_m^t \odot (I - y_m^t) \| / N, \quad (5)$$

where I is the identity matrix.

The smoothness term $P(\phi | I_m)$ or L_{smooth} in Eqn. 4 is computed over all voxels $p \in \Omega$ in the resampled I_m , where p is 1 when it is within the dimensions of I_m and 0 everywhere else. This term regularizes the deformations²⁷ and is computed by uniformly sampling the deformation field ϕ^t in Ω at each CLSTM step t as,

$$L_{smooth} = \sum_{t=1}^N \sum_{p \in \Omega} \|\nabla \phi^t(p)\|^2 / N. \quad (6)$$

Finally, rigidity or tumor preservation loss L_{tumor}^{pre} in Eqn. 4 penalizes any forward deformation flow within tumor voxels in moving image by requiring the Jacobian determinant $J(y) = \det(\nabla \phi(y))$ to be equal to 1, which corresponds to no mass change²⁸. Of note, this loss only constrains the deformation within the tumor voxels such that voxels outside the tumor geometry are allowed to deform. This loss is computed in N CLSTM steps as:

$$L_{tumor}^{pre} = \sum_{t=1}^N L_{tumor}^{pre,t} = \sum_{t=1}^N |\det(y_m^{t-1} \cdot \nabla \phi^t) - 1|_2 / N. \quad (7)$$

The formulation in Eqn. 4 is sufficient to achieve topology preserving deformations when no tumors are present in the fixed image. In order to account for any tumors occurring on the fixed image, similar to 4, we constrain the inverse deformation flow $\hat{\phi}$ as:

$$\begin{aligned} \log P(\hat{\phi} | I_m, y_m, I_f, y_f) &= \underbrace{\log P(I_m, y_m | I_f, y_f, \hat{\phi})}_{\text{Similarity}} + \underbrace{\log P(\hat{\phi} | I_f)}_{\text{Smoothness}} + \underbrace{\log P(\hat{\phi} | y_f)}_{\text{Rigidity}} + \text{constant}. \\ &= \hat{L}_{sim} + \hat{L}_{smooth} + L_{tumor}^{ob} \end{aligned} \quad (8)$$

The image similarity in Eqn. 8 is computed similar to Eqn. 5 using the resampled fixed image I_f and segmentation y_f at each CLSTM step t into the spatial coordinates of the moving image using $\hat{\phi}^t$. $\hat{\phi}^t$ is computed by the integration of negative flow field $\hat{\phi}^t = -\phi^t$ such that $\phi^t \circ \hat{\phi}^t = \exp(t) \circ \exp(-t) = \exp(t-t) = Id^8$. The approximation of inverse flow field as $-\phi$ is possible due to the small displacements resulting from a progressive deformation framework combined with a time integration performed inside each CLSTM²⁴. Smoothness loss is computed in the same way as in the forward case but with inverse DVF.

The rigidity penalty in Eqn. 8 is applied to tumor voxels occurring in the fixed image. This penalty is called tumor obliteration loss L_{tumor}^{ob} and it forces non-deformation of voxels on moving image that map to the voxels within the tumors occurring in the fixed image. The same loss is computed in 1 to N CLSTM steps as:

$$L_{tumor}^{ob} = \sum_{t=1}^N L_{tumor}^{ob,t} \sum_{t=1}^N \left| \det(y_f^{t-1} \cdot \nabla \hat{\phi}^t) - 1 \right|_2 / N, \quad (9)$$

The total loss is computed as,

$L_{reg} = L_{sim} + \hat{L}_{sim} + \lambda_{smooth} (L_{smooth} + \hat{L}_{smooth}) + \lambda_{pre} L_{tumor}^{pre} + \lambda_{ob} L_{tumor}^{ob} \cdot L_{sim}$ and $\hat{L}_{sim}$ are the image similarity loss; L_{smooth} and $\hat{L}_{smooth}$ are the smoothness loss and L_{tumor}^{pre} and L_{tumor}^{ob} are the tumor rigid constraint.

II.C. Training dataset and details

TRACER as well as compared DL-DIR methods were optimized in an unsupervised manner without any ground truth DVFs using identical training images and implemented using the Pytorch library to align 3DCTs from different patients. Contrast enhanced 3DCT image pairs ($n = 32,220$) from 180 patients with locally advanced non-small cell lung cancer (LA-NSCLC) treated with intensity modulated RT at our institution were used to train the models and then validated with 552 CT pairs from 24 patients not used in training to select the models for testing. All networks were trained on Nvidia A100 with 80GB memory and optimized using ADAM algorithm with an initial learning rate of 0.0002 for the first 50 epochs and then decayed to 0 in the next 50 epochs and a batch size of 4. We set $\lambda_{smooth} = 25$, $\lambda_{pre} = 1000$ and $\lambda_{ob} = 1000$ experimentally. Both TRACER and PACS used 8 3D-CLSTM steps. All models were “locked” for testing performed on entirely different datasets as described below and tuned to the best of our ability in order to achieve the best performance possible.

Preprocessing details:

All CT scans were cropped to contain the chest region with 10 cm margin added to include a small portion of air outside the body such that both lungs and the chest enclosing the lungs were visible in all scans and then resampled to $128 \times 128 \times 96$ volume. Images were resampled back to the original volumes for accuracy calculation.

II.D. Testing datasets

Dataset I consisting of 380 image pairs (20×19) sourced from a public dataset²⁹ of pre-RT thoracic CT scans of LA-NSCLC patients (tumor size between 4.88 cc to 356 cc) was used to evaluate registration accuracy when using auto-segmented tumors²² as conditioning inputs. Expert tumor delineations were available and used for accuracy evaluation.

Dataset II consisting of 756 CT image pairs (28×27) from a randomly selected set of 28 LA-NSCLC patients with expert delineations from public dataset³⁰ and treated with definitive IMRT were evaluated. Tumors ranged in size between 0.022 cc to 640 cc and were distributed across the various lobes of the lungs including the mediastinum.

Dataset III consisting of 42 institutional LA-NSCLC patients (23 male and 19 female) treated with IMRT to 60 Gy in 2 Gy fractions were aligned to a canonical reference to study

utility of TRACER for VBA. Retrospective analysis was approved by the local institutional review board at the Memorial Sloan Kettering Cancer Center.

II.E. Metrics and statistical analysis

Registration accuracy was evaluated in terms of healthy tissue segmentation accuracy computed using the Dice similarity coefficient (DSC) and Hausdroff distance at 95th percentile (HD95) metrics for the lungs, heart, and spinal cord by comparing against deep learning model³¹ used to generate clinical segmentations. Instead of measuring target registration error, which requires identification of matching landmarks that is hard to locate on two different patient anatomies, we measured the distance between the medial axes of tubular structures including, the pulmonary artery (PA), aorta, inferior venacava (IVC), and trachea from a subset of 26 patients with available delineations. This distance, called median of closest points distance (MCD) measured misalignments along the entire organ length.

Accuracy of tumor preservation was computed using tumor volume loss or

$\Delta T = \frac{|V_{def} - V_m|}{V_m} \times 100$, where V_m, V_{def} are the volume of tumors in the original and deformed moving images, respectively. The percentage mean local expansion and shrinkage inside the tumor in the moving image ($M_{tex} \%$) was calculated using the Jacobian determinant as $\frac{\sum \|J_i - 1\| \cdot y_m}{N} \times 100$ ³², where N is the number of voxels within the tumor mask in the original

moving image. We also evaluated the tumor preservation by computing the mean squared error (MSE) between the resampled tumor mask within the resampled moving image and the original tumor mask within the moving image. In a subset of 42 patients (Dataset III) who had radiation dose maps available, we also computed difference in the Planned Tumor Dose (ΔPTD) between undeformed and deformed moving images. Registration accuracy was computed by comparing the resampled lung contour from moving patient with the lung contour of the reference patient using DSC overlap measure. A DSC < 0.8 was deemed as poor registration. The number of patients with poor registration were calculated. Statistical comparisons of TRACER against the various methods was performed using two-sided, paired Wilcoxon signed-rank tests at 95% confidence interval.

III. Experiments configuration

III.A. Comparative experiments

TRACER (1,026,771 parameters) was compared against PACS²¹ (522,723 parameters) model as the two models share similar encoder-decoder structure. It was also compared against fast symmetric registration (FastSyM)¹⁷, which performs bidirectional alignment to compute diffeomorphic registration, transformer based Transmorph¹², and iterative DIR called SyN³³.

III.B. Ablation experiments

Ablation experiments were done using Dataset II. Ablation experiments studied the impact of various losses and CLSTM on tumor volume preservation and normal tissue segmentation accuracy. Impact of number of CLSTM layers was also analyzed in the context of strength of encoder feature activation and segmentation accuracy of the heart as it occurs centrally.

III.C. Results

III.C.1. Registration accuracy

Table 1 shows the tumor preservation metrics computed from the three datasets. TRACER best preserved tumor volumes as shown by the lowest tumor loss ΔT and shrinkage M_{loss} metrics in all testing datasets. It also resulted in the least MSE indicating that the CT intensities within the moving tumor image were preserved upon deformation. PACS on the other hand was least effective in preserving tumor, when using local tumor expansion and shrinkage measure. PACS was similarly accurate as the Transmorph method when using MSE measure. Both FastSym¹⁷ and transformer-based Transmorph¹² were less accurate than TRACER, indicating the importance of tumor conditioning.

In terms of aligning healthy tissues, PACS was the most accurate method followed by TRACER (Figure. 3). Transmorph showed wide variation in accuracy for the lungs and heart with HD95 metric. FastSym was slightly less accurate than TRACER. Improved accuracy with PACS for normal tissues was also evident when computing the MCD metric as shown in Table. 2. TRACER was less accurate than PACS for all tubular structures except trachea. All DL methods resulted in a mean MCD that was within the diameter of aorta (ranging from 3.3 cm for females to 3.6 cm for males)³⁴ and the trachea 2.3 cm on average³⁵. Only PACS, TRACER and FastSym resulted in a mean MCD within the known diameter variations of the PA of 2.7 cm to 3.3 cm³⁶. PACS, TRACER, and Transmorph achieved mean MCD within acceptable diameter of IVC in healthy and non-hypertensive patients ($1.45 \pm .16$ cm)³⁷. Hypertensive patients have a larger diameter of IVC. SyN was the least accurate of all the methods. Fig. 4 shows the medial axis skeleton from clinical delineation on the fixed image for a reference patient together with the medial axis skeletons produced from deformed segmentations produced using PACS and TRACER.

III.C.2. Qualitative results

Fig.5 shows registration results for two representative examples with varying tumor location and sizes. The second case also shows a female aligned to a male patient. As shown, TRACER best preserved the tumor occurring in the moving image and did not produce unrealistic distortions in places containing tumor on the fixed image. PACS resulted in complete erosion of the tumor in the moving image as well as unrealistic deformations in the location corresponding to tumor occurring in the fixed image. FastSym was slightly more robust to the presence of tumors but was not able to preserve the tumor on the moving image as well as TRACER. Both PACS and Transmorph produced unrealistic stretching of the right proximal bronchial tree to a considerable extent in the second patient and to a lesser degree in the first patient due a tumor on the fixed image abutting the right mediastinum. The distortion of the pulmonary artery is also visible for both patients when using PACS and Transmorph. TRACER on the other hand was unaffected by the tumor on the fixed image and preserved the tumor on the moving images.

III.C.3. Suitability for VBA

Registration quality filtering was performed using DSC metric by comparing resampled and reference lung contours (lung DSC < 0.8) similar to registration filtering performed in prior

work^{5,6}. The count of patients excluded through such filtering is shown in Table. 3. Our analysis showed that both PACS and Transmorph resulted in the least number of excluded patients followed by TRACER. SyN³³ that is commonly used in VBA analysis resulted in an exclusion of $\geq 45\%$ of patients. FastSym was less effective than TRACER but more accurate SyN. TRACER showed similar variation in the patients that were removed for the two different references, indicating robustness to gender and lung lobes.

On the other hand, PACS resulted in the largest error in the planned dose to the tumor measured between undeformed moving and deformed moving image, indicating lack of tumor geometry preservation. FastSym and Transmorph were similar in terms of the tumor dose difference. Although SyN showed a small error in the dose to the tumor, it was ineffective in aligning most patients. TRACER resulted in the least planned tumor dose error.

III.C.4. Ablation experiments

Table 4 shows the tumor preservation and tissue segmentation accuracies with and without CLSTM and with and without tumor conditioning constraints. Removal of tumor conditioning drastically reduced the tumor preservation accuracy compared to when using these constraints. Tumor conditioning in the forward direction was more important for tumor preservation than in the inverse direction. Conversely, tumor conditioning reduced the segmentation accuracy for the healthy tissues, which was slightly improved when using CLSTM. On the other hand, CLSTM without tumor conditioning reduced tumor preservation accuracy furthermore as the model tries to progressively shrink the tumor voxels to better align with the healthy tissue occurring in the same anatomic location. Fig. 6 shows an example of progressively refined alignment without and with tumor conditioning through the different CLSTM steps. As seen, without the CLSTM, the tumor geometry is progressively lost in the former case but is preserved in the latter case. On the other hand, when not using tumor conditioning, the chest wall is pulled towards lungs to match with the tumor occurring in the fixed image, thus resulting in physically unrealistic deformation compared to when using the tumor conditioning with CLSTM. The successive deformation maps are also shown alongside the deformed images.

Next, we studied the importance of using a stacked CLSTM on the registration accuracy. As shown in Fig. 7, the feature activations are more pronounced towards voxels with large differences between the two patient scans such as the heart and the tumor when using the CLSTM. Furthermore, different areas of the image show higher feature activations for different CLSTM steps, which allows TRACER to capture deformations at varying scales and locations. Feature activations across the encoder layers follow different trends across the steps when using CLSTM with the activations decreasing drastically in the later steps for layer 1 and increasing for layer 3 as shown in Fig. 7 C. The feature activations are relatively low and follow the same trend when not using the stacked CLSTM. Also, the registration accuracy increases with the number of stacked CLSTMs (Fig. 7 D).

IV. Discussion

We introduced TRACER, an inter-patient DL-DIR method for aligning thoracic CTs of patients with LA-NSCLC. TRACER was the most accurate approach for preserving conditioned structures such as the tumors and prevented physically unrealistic anatomy deformations in locations where new tumors occurred on fixed image. TRACER was second most accurate for aligning organs and tubular organs after PACS²¹. PACS on the other hand was ineffective to preserve tumors. Unlike prior approaches using rigidity constraints^{5,15}, TRACER is fully automated as it does not require manual delineations for conditioned structures and can handle tumors occurring at non-corresponding locations in moving and fixed images.

VBA methods have primarily utilized iterative non deep learning methods including rigid registration⁷ and elastic DIR^{1,2,5}. We, for the first time analyzed the feasibility of using DL-DIR methods for addressing inter-patient tumor variabilities for VBA. Our analysis showed that TRACER as well as other DL-DIR methods were more accurate than iterative DIR using SyN³³. Despite preserving tumor geometry, SyN resulted in the largest number of excluded examples due to poor registration. TRACER on the other hand was resilient to anatomic differences even in the presence of non-corresponding tumors and patient gender as evidenced by relatively fewer excluded cases with highest tumor geometry preservation. Of note, we used lung registration accuracy as a metric to filter patients, but structures can also be used. TRACER was resistant to the presence of tumors abutting mediastinum and better preserved geometry compared with PACS and Transmorph.

TRACER shares the same registration network architecture as PACS²¹. TRACER does not include the segmentation subnetwork used in PACS, and makes use of explicit tumor conditioning as well as bidirectional tumor incompressibility constraints. Our results showed improved tumor preservation when combining tumor conditioning with the stacked 3DCLSTM approach. In particular, removal of tumor conditioning resulted in a dramatic tumor mass loss, thus suggesting that conditioning is critical for tumor preservation and removing physically unrealistic deformations. A key advantage of the stacked CLSTM architecture is that it allows to capture varying amounts of local tissue deformation in the different layers. In addition, CLSTM naturally allows to compute spatially and temporally varying velocity field, which is not possible with single step registration methods^{17,25}. It is known that the ability to model spatially and temporally varying velocity fields reduces chances of high energy transformations in the presence of substantial anatomic differences²⁴.

As limitation, the current work was focused on developing and evaluating a inter-patient registration method and hence did not perform an indepth study of downstream effects on VBA itself, which is for future studies. In summary, TRACER generated tumor preserving accurate registration of inter-patient thoracic CTs.

V. Conclusion

We introduced TRACER, an inter-patient registration method for aligning thoracic CTs of patients with LA-NSCLC. TRACER was the most accurate approach for preserving conditioned structures such as the tumors and prevented physically unrealistic anatomy deformations in locations where new tumors occurred on fixed image. TRACER also showed capability to balance registration accuracy for normal tissues while preserving tumor geometry, thus indicating its suitability for voxel-based analysis. Further studies are required to evaluate its utility in population level outcomes modeling studies using voxel-based analysis.

Acknowledgments

This work was partially supported by NCI R01CA25881 and the MSK Cancer Center support grant/core grant P30 CA008748, and the Korea Health Technology R&D Project through the Korea Health Industry Development Institute (KHIDI), funded by the Ministry of Health Welfare, Republic of Korea (HI19C1234).

This work was partially supported by the NCI R01CA258821 and the Korea Health Technology R&D Project HI19C1234 and the MSK Cancer Center core grant P30 CA008748.

References

1. Palma G, Monti S, D'Avino V, Conson M, Liuzzi R, Pressello M, Donato V, Deasy J, Quarantelli M, Pacelli R, and Cella L, "A voxel-based approach to explore local dose differences associated with radiation-induced lung damage," *Int J. Radiat Oncol Biol Phys*, vol. 96, no. 1, pp. 127–33, 2016. [PubMed: 27511851]
2. Craddock M, Nestle U, Koenig J, Schimek-Jasch T, Kremp S, Lenz S, Banfill K, Davey A, Price G, Salem A, Faivre-Finn C, van Herk M, and McWilliam A, "Cardiac function modifies the impact of heart base dose on survival: A voxel-wise analysis of patients with lung cancer from the PET-plan trial," *J Thorac Oncol*, vol. 18, no. 1, pp. 57–66, 2023. [PubMed: 36130693]
3. Beasley W, Thor M, McWilliam A, Green A, M. R, Slevin N, Olsson C, Pettersson N, Finizia C, Estilo C, Riaz N, Lee N, Deasy J, and van Herk M, "Image-based data mining to probe dosimetric correlates of radiation-induced trismus," *Int J Radiat Oncol Biol Phys*, vol. 102, no. 4, pp. 1330–1338, 2019.
4. Monti S, Pacelli R, Cella L, and Palma G, "Inter-patient image registration algorithms to disentangle regional dose bioeffects," *Sci Rep*, vol. 8, p. 4915, 2018. [PubMed: 29559687]
5. Mylona E, Acosta O, Lizée T, Lafond C, Crehange G, Magné N, Chiavassa S, Supiot S, Ospina Arango JD, Campillo-Gimenez B, Castelli J, and de Crevoisier R, "Voxel-based analysis for identification of urethrovessical subregions predicting urinary toxicity after prostate cancer radiation therapy," *Int J Radiat Oncol Biol Phys*, vol. 104, no. 2, pp. 343–354, 2019. [PubMed: 30716523]
6. Gil N and Lipton FR, M.L, "Registration quality filtering improves robustness of voxel-wise analyses to the choice of brain template," *Neuroimage*, vol. 227, p. 117657, 2021.
7. Beaumont J, Acosta O, Devillers A, Palard-Nevello X, Chajon E, Crevoisier R. d., and Castelli J, "Voxel-based identification of local recurrence sub-regions from pretreatment pet/ct for locally advanced head and neck cancers," *EJNMMI Res*, vol. 9, no. 90, 2019.
8. Dalca AV, Balakrishnan G, Guttag J, and Sabuncu MR, "Unsupervised learning of probabilistic diffeomorphic registration for images and surfaces," *Medical Image Anal.*, vol. 57, pp. 226–236, 2019.
9. Xu Z and Niethammer M, "Deepatlas: Joint semi-supervised learning of image registration and segmentation," in *MICCAI*. Springer, 2019, pp. 420–429.
10. Haq R, Berry SL, Deasy J, Hunt M, and Veeraraghavan H, "Dynamic multiatlas selection-based consensus segmentation of head and neck structures from ct images," *Med Phys*, vol. 46, no. 12, pp. 5612–5622, 2019. [PubMed: 31587300]

11. Liu F, Yan K, Harrison AP, Guo D, Lu L, Yuille AL, Huang L, Xie G, Xiao J, Ye X et al. , "SAME: Deformable image registration based on self-supervised anatomical embeddings," in Medical Image Computing and Computer-Assisted Intervention, 2021, pp. 87–97.
12. Chen J, Frey EC, He Y, Segars WP, Li Y, and Du Y, "Transmorph: Transformer for unsupervised medical image registration," Medical image analysis, vol. 82, p. 102615, 2022.
13. Ahmad S, Fan J, Dong P, Cao X, Yap P-T, and Shen D, "Deep learning deformation initialization for rapid groupwise registration of inhomogeneous image populations," Front Neuroinform, vol. 34, no. 13, 2019.
14. Zhang Y, Wu X, Gach H, Li H, and Yang D, "Groupregnet: a groupwise one-shot deep learning-based 4d image registration method," Phys Med Biol, vol. 66, no. 4, p. 045030. [PubMed: 33412539]
15. Staring M, Klein S, and Pluim J, "A rigidity penalty term for nonrigid registration," Medical Physics, vol. 34, no. 11, pp. 4098–108, 2007. [PubMed: 18072476]
16. König L, Derksen A, Papenberg N, and Haas B, "Deformable image registration for adaptive radiotherapy with guaranteed local rigidity constraints," Radiat Oncol, vol. 11, no. 122, 2016.
17. Mok TCW and Chung ACS, "Large deformation diffeomorphic image registration with laplacian pyramid networks," in Medical Image Computing and Computer Assisted Intervention. Springer International Publishing, 2020, pp. 211–221.
18. "Cyclemorph: Cycle consistent unsupervised deformable image registration," Medical Image Analysis, vol. 71, p. 102036, 2021. [PubMed: 33827038]
19. Zhang J, "Inverse-consistent deep networks for unsupervised deformable image registration," CoRR, vol. abs/1809.03443, 2018.
20. Sandkühler R, Andermatt S, Bauman G, Nyilas S, Jud C, and Cattin PC, "Recurrent registration neural networks for deformable image registration," Adv. Neural Inf. Process. Syst, vol. 32, 2019.
21. Jiang J and Veeraraghavan H, "One shot PACS: Patient specific anatomic context and shape prior aware recurrent registration-segmentation of longitudinal thoracic cone beam CTs," IEEE Trans Med Imaging, 2022.
22. Jiang J, Hu Y.-c., Liu C-J, Halpenny D, Hellmann MD, Deasy JO, Mageras G, and Veeraraghavan H, "Multiple resolution residually connected feature streams for automatic lung tumor segmentation from ct images," IEEE Trans Med Imaging, vol. 38, no. 1, pp. 134–144, 2018. [PubMed: 30040632]
23. Jaderberg M, Simonyan K, Zisserman A, and Kavukcuoglu K, "Spatial transformer networks," arXiv preprint arXiv:1506.02025, 2015.
24. Ashburner J, "A fast diffeomorphic image registration algorithm," Neuroimage, vol. 38, no. 1, pp. 95–113. [PubMed: 17761438]
25. Balakrishnan G, Zhao A, Sabuncu MR, Guttag J, and Dalca AV, "Voxelmorph: a learning framework for deformable medical image registration," IEEE transactions on medical imaging, vol. 38, no. 8, pp. 1788–1800, 2019.
26. Yang Y and Soatto S, "Conditional prior networks for optical flow," in Proceedings of the European Conference on Computer Vision (ECCV), 2018, pp. 271–287.
27. Mok TCW and Chung ACS, "Conditional deformable image registration with convolutional neural network," in Medical Image Computing and Computer Assisted Intervention – MICCAI 2021. Springer International Publishing, 2021, pp. 35–45.
28. Niedzielski J, Yang J, Stingo F, Martel M, Mohan R, Gomez D, Briere T, Liao Z, and Court L, "Objectively quantifying radiation esophagitis with novel computed tomography-based metrics," Intl J Radiat Oncol Bio Physics, vol. 94, no. 2, pp. 385–93, 2016.
29. Kalendralis P, Shi Z, Traverso A, Choudhury A, Sloep M, Zhovannik I, Starmans MP, Grittner D, Feltens P, Monshouwer R et al. , "FAIR-compliant clinical, radiomics and DICOM metadata of RIDER, interobserver, lung1 and head-neck1 TCIA collections," Med Phys, vol. 47, no. 11, pp. 5931–5940, 2020. [PubMed: 32521049]
30. "Data from NSCLC-radiomics. The Cancer Imaging Archive," 2015.
31. Um H, Jiang J, Thor M, Rimner A, Luo L, Deasy JO, and Veeraraghavan H, "Multiple resolution residual network for automatic thoracic organs-at-risk segmentation from ct," 2020.

32. Alam SR, Zhang P, Zhang S-Y, Chen I, Rimner A, Tyagi N, Hu Y-C, Lu W, Yorke ED, Deasy JO, and Thor M, "Early prediction of acute esophagitis for adaptive radiation therapy," *Int J Radiat Oncol Biol Phys*, vol. 110, no. 3, pp. 883–892, 2021. [PubMed: 33453309]
33. Avants BB, Epstein CL, Grossman M, and Gee JC, "Symmetric diffeomorphic image registration with cross-correlation: evaluating automated labeling of elderly and neurodegenerative brain," *Medical image analysis*, vol. 12, no. 1, pp. 26–41, 2008. [PubMed: 17659998]
34. Mao S, Ahmadi N, Shah B, Beckmann D, Chen A, Ngo L, Flores F, Gao Y, and Budoff M, "Normal thoracic aorta diameter on cardiac computed tomography in healthy asymptomatic adult; impact of age and gender," *Acad Radiol*, vol. 15, no. 7, pp. 827–834, 2008. [PubMed: 18572117]
35. Epstein S, "Anatomy and physiology of tracheostomy," *Respir Care*, vol. 50, no. 4, pp. 476–82, 2005. [PubMed: 15807905]
36. Edwards P, Bull R, and Coulden R, "CT measurement of main pulmonary artery diameter," *Br J Radiol*, vol. 71, no. 850, pp. 1018–20, 1998. [PubMed: 10211060]
37. Besli F, Kecebas M, Caliskan S, Dereli S, Baran I, and Turker Y, "The utility of inferior vena cava diameter and the degree of inspiratory collapse in patients with systolic heart failure," *Am J Emer Med*, vol. 33, no. 5, pp. 653–7.
38. Jiang J and Veeraraghavan H, "One shot PACS: Patient specific anatomic context and shape prior aware recurrent registration-segmentation of longitudinal thoracic cone beam CTs," *IEEE Trans Med Imaging*, p. PP:10.1109/TMI.2022.3154934, 2022.

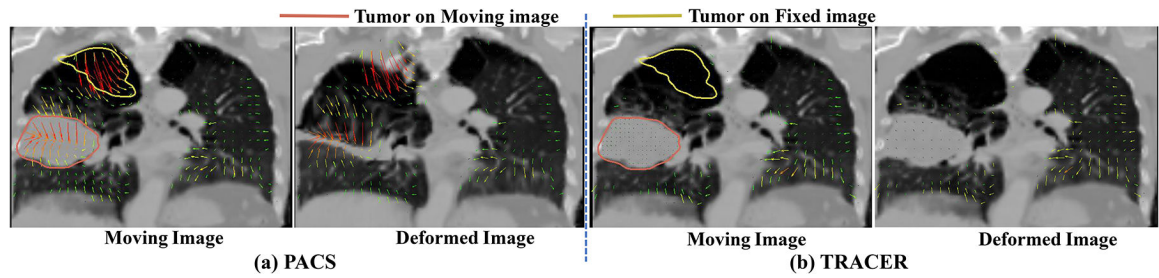
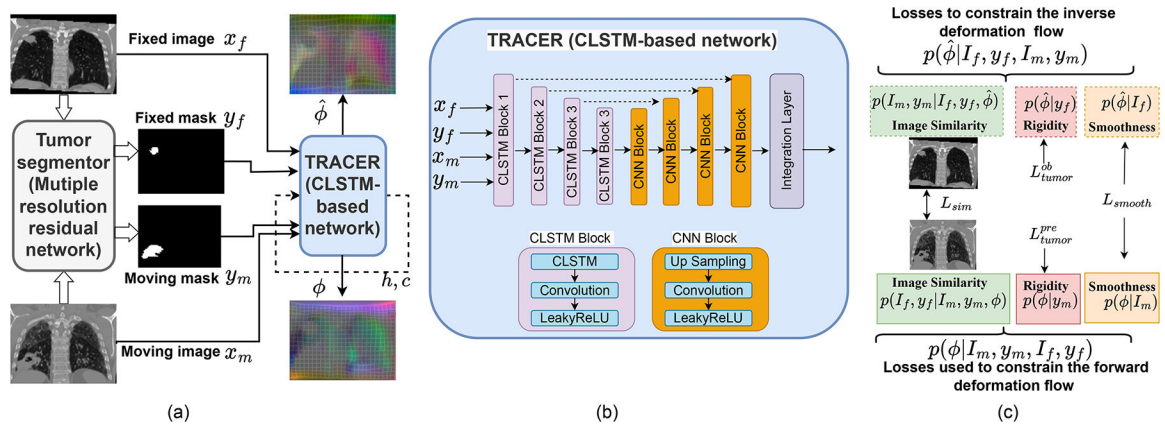


Figure 1:

Rationale for tumor conditioned inter-patient registration. (a) Lack of tumor conditioning in PACS, ill-preserves tumors and produces physically unrealistic stretching of tissues. (b) Tumor conditioning in TRACER preserves tumor and normal tissue topology.

**Figure 2:**

(a) Schematic of tumor aware recurrent inter-patient DIR (TRACER) showing the fixed image and its mask ($\{x_f, y_f\}$), moving image and its mask ($\{x_m, y_m\}$) and the hidden layer h input to the TRACER. Tumor conditioning is automated using published method²². (b) shows TRACER architecture. (c) shows the losses used to optimize the network.

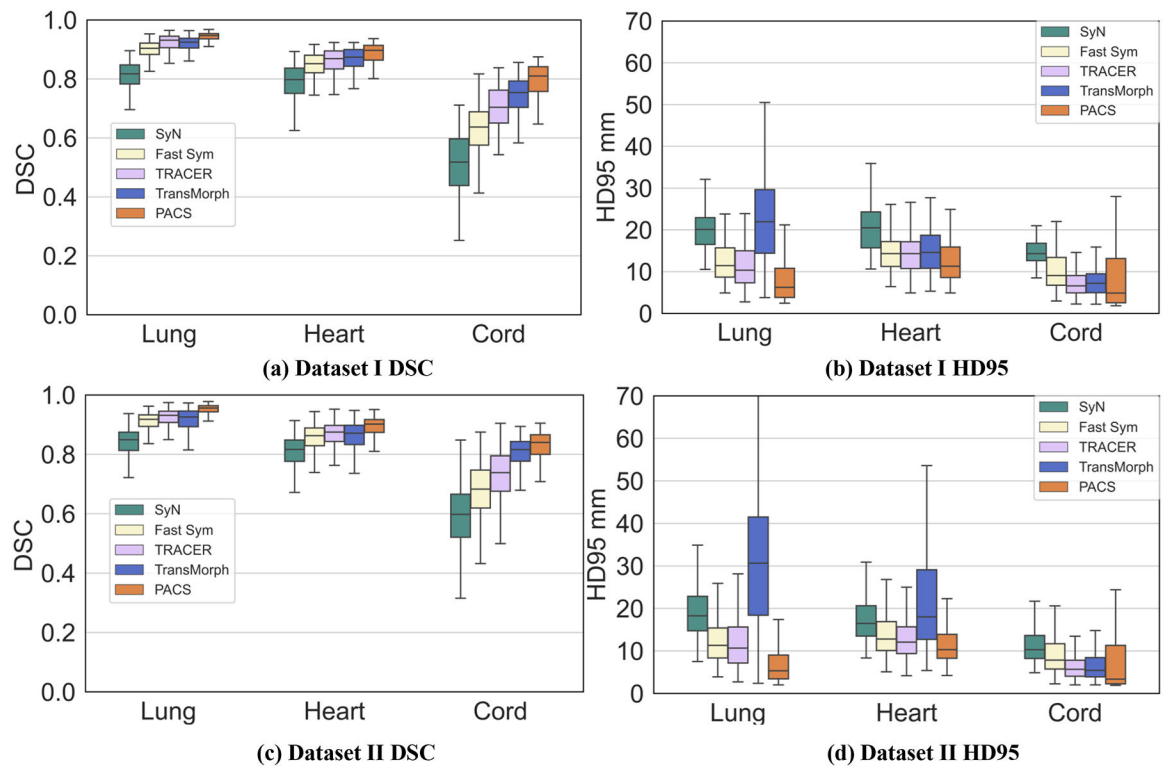


Figure 3:
DSC and HD95 accuracies of various methods on the testing datasets.

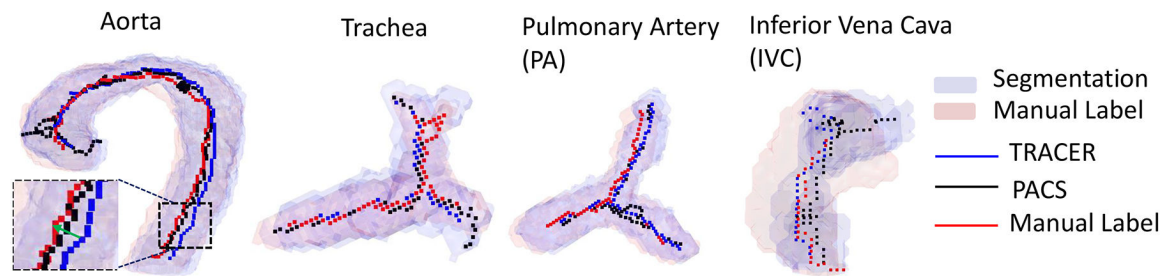


Figure 4:
Medial axis skeletons produced from PACS, TRACER, and manual segmentation of aorta, PA, IVC, and trachea on a sample patient.

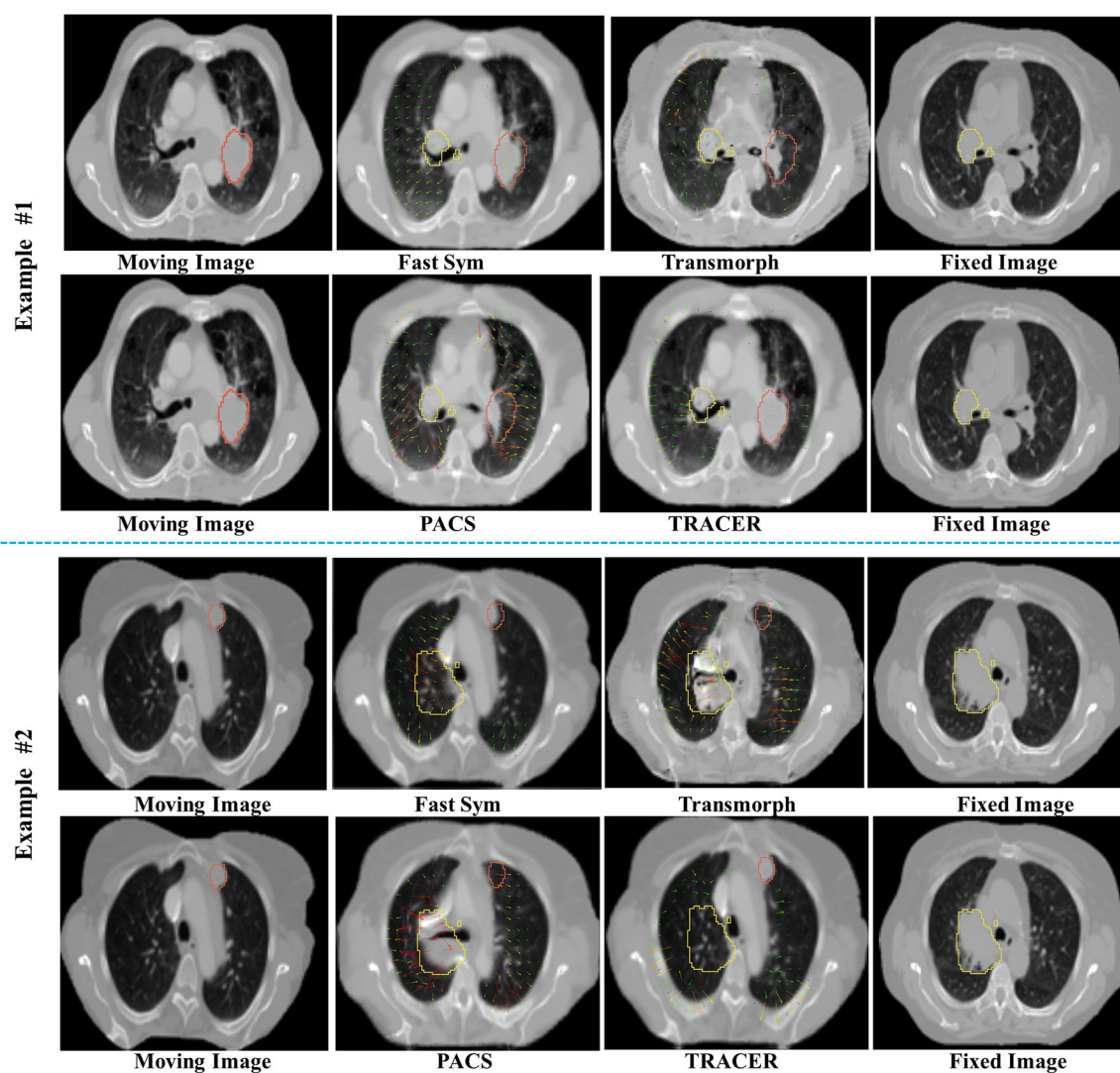


Figure 5: Registration for 2 different patient pairs containing tumor on completely different locations as well as different gender (Example #2). Red and yellow contours indicate the tumor region on moving and fixed images.

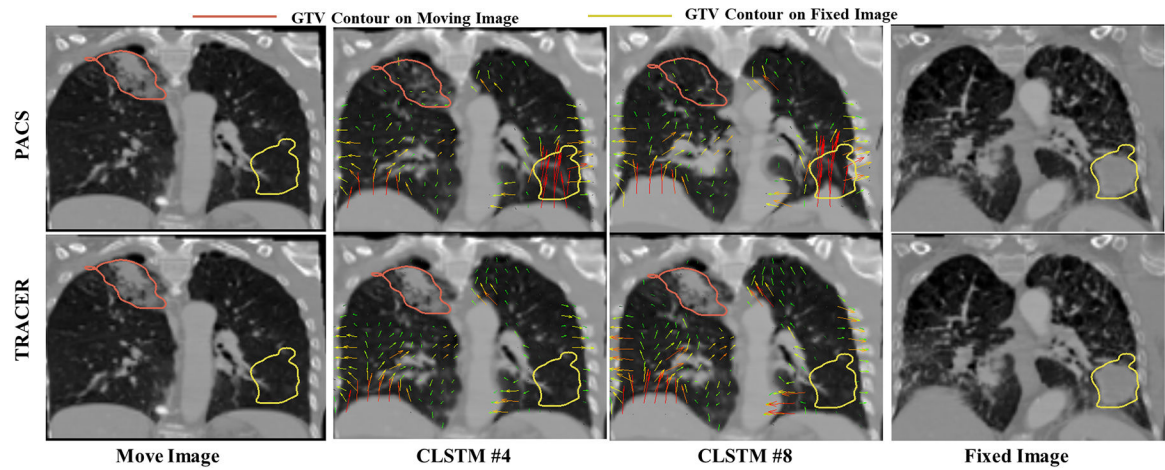


Figure 6:

TRACER used to align a representative pair of images containing large tumors on different lobes of the lungs. Deformation map (top row) and resampled moving images (bottom row) produced in CLSTM steps 2,4,6, and 8 (b) without tumor preservation (upper) and with tumor preservation constraints (below) are also shown.

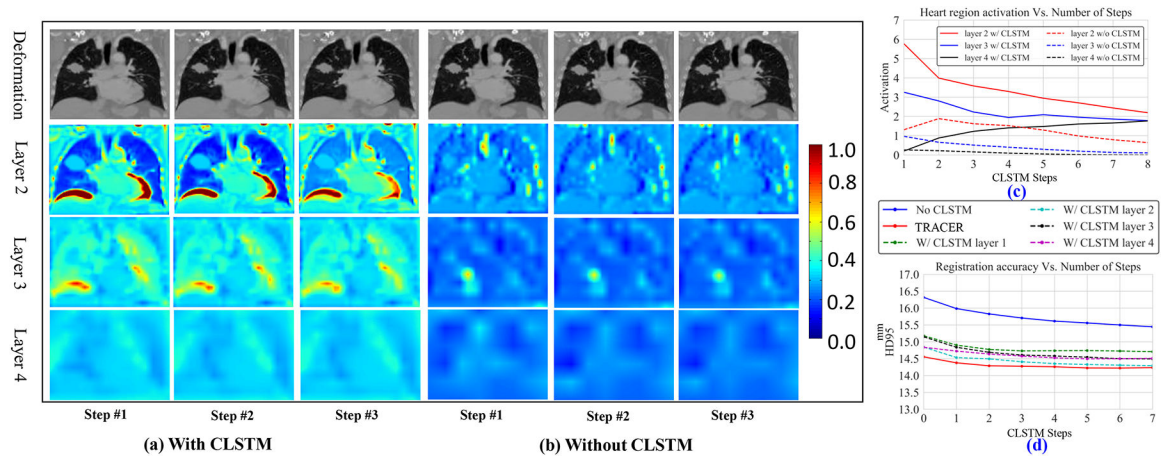


Figure 7:

(b) Feature activations produced in CNN layers 2, 3, and 4 for steps 1, 2, 3 without CLSTM, and (a) with CLSTM. (c) shows mean feature activations within heart in the layers 2, 3, and 4. (d) shows DSC accuracy with increasing number of CLSTM steps.

Table 1:

Tumor preservation metrics computed on three testing datasets using the metrics of ΔT , M_{lexs} % and MSE.

Method	Dataset I (380 pairs)			Dataset II (756 pairs)			Dataset III (41 pairs)		
	ΔT % ↓	M_{lexs} % ↓	MSE ↓	ΔT % ↓	M_{lexs} % ↓	MSE ↓	ΔT % ↓	M_{lexs} % ↓	MSE ↓
SyN ³³	6.91±4.73	11.11±3.50	0.14±0.076	8.74±5.62	11.0±4.85	0.064±0.017	3.18±3.42	3.16±1.85	0.040±0.048
Fast Sym ¹⁷	18.87±13.67	14.0±5.23	0.21±0.057	18.19±31.03	15.0±7.23	0.25±0.12	17.36±13.69	3.89±2.47	0.12±0.062
Transmorph ¹²	54.51±22.18	23.0±7.18	0.52±0.093	31.12±57.16	27.0±10.03	0.45±0.15	45.74±27.66	4.27±2.26	0.37±0.19
PACS ³⁸	96.48±23.48	33.13±6.27	0.51±0.097	58.98±92.10	29.0±11.33	0.47±0.15	97.93±96.68	7.36±2.45	0.37±0.20
TRACER	0.24±0.49	0.8±0.09	0.005±0.003	0.40±0.82	0.45±0.12	0.005±0.006	0.13±0.15	0.87±0.14	0.003±0.002

Table 2:

Accuracy of aligning medial axis skeleton (MAS) measured as mean of median of closest points for tubular structures.

Method	52 pairs from different patients			
	Aorta (cm) ↓	Trachea (cm) ↓	PA (cm) ↓	IVC (cm) ↓
SyN ³³	3.70±0.81	2.31±0.20	4.82±2.79	4.95±0.65
Fast Sym ¹⁷	2.56±0.77	0.38±0.18	3.28±2.15	1.78±0.54
Transmorph ¹²	1.95±0.75	0.40±0.25	3.48±2.31	1.02±0.58
PACS ³⁸	1.06±0.71	0.15±0.15	1.98±1.64	0.65±0.45
TRACER	1.59±0.76	0.13±0.14	2.49±1.93	0.79±0.53

Table 3:

Suitability for VBA. Number of excluded patients (lung DSC < 0.8) and tumor dose preservation Δ PTD Gy by various methods measured using 41 cases aligned to canonical female and male reference in Dataset III.

Method	Female reference ↓			Male reference ↓		
	Excluded cases		Tumor dose difference	Excluded cases		Tumor dose difference
	Left Lung	Right Lung	Δ PTD Gy	Left Lung	Right Lung	Δ PTD Gy
SyNs ³³	18	26	0.01±0.01	21	26	0.01±0.01
Fast Sym ¹⁷	11	11	0.17±0.38	7	16	0.41±0.75
Transmorph ¹²	0	0	0.14±0.14	1	2	0.24±0.26
PACS ³⁸	0	0	28.45±15.30	1	2	22.92±19.91
TRACER	5	5	0.01±0.02	6	5	0.013±0.027

Table 4:

Ablation testing. Forward conditioning uses moving image, inverse uses fixed image, and both uses both images.

CLSTM	Condition	ΔT % ↓	Lung ↑	Heart ↑	Cord ↑
×	None	80.2±24.1	0.94±0.02	0.86±0.05	0.75±0.09
×	Inverse	30.8±18.9	0.93±0.02	0.85±0.05	0.69±0.09
×	Forward	0.83±1.56	0.93±0.02	0.85±0.05	0.69±0.09
×	Both	0.54±0.67	0.90±0.06	0.85±0.06	0.68±0.10
✓	None	96.5±23.5	0.94±0.02	0.87±0.05	0.76±0.09
✓	Inverse	41.9±46.7	0.93±0.03	0.85±0.05	0.72±0.09
✓	Forward	0.31±0.61	0.93±0.03	0.85±0.05	0.72±0.10
✓	Both	0.24±0.49	0.92±0.03	0.85±0.06	0.70±0.09