

# LungRegNet: An unsupervised deformable image registration method for 4D-CT lung

Yabo Fu\*, Yang Lei\*, Tonghe Wang, Kristin Higgins, Jeffrey D. Bradley, Walter J. Curran, Tian Liu, and Xiaofeng Yang<sup>a)</sup>

*Department of Radiation Oncology and Winship Cancer Institute, Emory University, Atlanta, GA 30322, USA*

(Received 23 October 2019; revised 9 January 2020; accepted for publication 27 January 2020; published 26 February 2020)

**Purpose:** To develop an accurate and fast deformable image registration (DIR) method for four-dimensional computed tomography (4D-CT) lung images. Deep learning-based methods have the potential to quickly predict the deformation vector field (DVF) in a few forward predictions. We have developed an unsupervised deep learning method for 4D-CT lung DIR with excellent performances in terms of registration accuracies, robustness, and computational speed.

**Methods:** A fast and accurate 4D-CT lung DIR method, namely LungRegNet, was proposed using deep learning. LungRegNet consists of two subnetworks which are CoarseNet and FineNet. As the name suggests, CoarseNet predicts large lung motion on a coarse scale image while FineNet predicts local lung motion on a fine scale image. Both the CoarseNet and FineNet include a generator and a discriminator. The generator was trained to directly predict the DVF to deform the moving image. The discriminator was trained to distinguish the deformed images from the original images. CoarseNet was first trained to deform the moving images. The deformed images were then used by the FineNet for FineNet training. To increase the registration accuracy of the LungRegNet, we generated vessel-enhanced images by generating pulmonary vasculature probability maps prior to the network prediction.

**Results:** We performed fivefold cross validation on ten 4D-CT datasets from our department. To compare with other methods, we also tested our method using separate 10 DIRLAB datasets that provide 300 manual landmark pairs per case for target registration error (TRE) calculation. Our results suggested that LungRegNet has achieved better registration accuracy in terms of TRE than other deep learning-based methods available in the literature on DIRLAB datasets. Compared to conventional DIR methods, LungRegNet could generate comparable registration accuracy with TRE smaller than 2 mm. The integration of both the discriminator and pulmonary vessel enhancements into the network was crucial to obtain high registration accuracy for 4D-CT lung DIR. The mean and standard deviation of TRE were  $1.00 \pm 0.53$  mm and  $1.59 \pm 1.58$  mm on our datasets and DIRLAB datasets respectively.

**Conclusions:** An unsupervised deep learning-based method has been developed to rapidly and accurately register 4D-CT lung images. LungRegNet has outperformed its deep-learning-based peers and achieved excellent registration accuracy in terms of TRE. © 2020 American Association of Physicists in Medicine [<https://doi.org/10.1002/mp.14065>]

Key words: 4D-CT lung, deep learning, deformable image registration, unsupervised learning

## 1. INTRODUCTION

Deformable image registration (DIR) of four-dimensional computed tomography (4D-CT) lung images is important in multiple radiation therapy applications including lung motion tracking,<sup>1</sup> target definition,<sup>2</sup> image fusion,<sup>3</sup> gated treatment planning<sup>4</sup> and treatment response evaluations.<sup>5</sup> Though DIR of 4D-CT lung has been extensively studied over the past few decades,<sup>6,7</sup> it remains challenging to accurately and quickly register 4D-CT lung images due to the large lung motion and bulky image sizes. To meet the increasingly demanding medical needs, it is necessary to develop a DIR method that can achieve fast computational speed and high registration accuracy simultaneously.

Four-dimensional computed tomography lung scan has been widely used in radiation therapy to aid treatment

planning to spare healthy tissue and increase dose delivery to the tumor target. Currently, respiratory gating for lung stereotactic body radiation therapy (SBRT) is regularly performed using varian real-time position management (RPM) system.<sup>8</sup> The RPM system relies on precise correlation between camera-captured upper abdominal motion and the actual lung tumor pattern. Fast lung motion tracking could facilitate online individualized correlation establishment between the camera-captured upper abdominal motion and the patient-specific lung motion pattern. Accurate lung motion modeling allows respiratory phase-based planning target volume (PTV) and organ-at-risk (OAR) dose calculation. Quantitative phase-based dose calculation would enable us to choose optimal gating phases based on the calculated PTV/OAR dose volume histograms. DIR of 4D-CT lung is a viable solution for fast and accurate lung motion tracking and modeling.



Conventional image similarity measures such as sum of squared differences (SSD) and mean absolute difference (MAE) used in DIR methods such as optical flow<sup>9–11</sup> and demons<sup>12,13</sup> are dependent on the assumption that image intensities are consistent between the fixed and moving images. However, image intensity of 4D-CT lung changes as air density changes within the lung throughout a respiratory cycle. As a result, intensity compensation is necessary for accurate lung registration using intensity-based DIR methods.<sup>14</sup> To address the issue, normalized cross correlation (NCC) and mutual information (MI) are often used to measure image similarity. Conventional DIR methods are generally iterative and slow especially for large 4D-CT datasets. In addition, iterative methods are mostly susceptible to local minima.<sup>6,7</sup> DIR is an ill-posed problem since there are multiple possible deformation vector fields (DVF) that can deform the moving image to match the fixed image. DVF regularization is necessary to generate plausible lung motion. Conventional methods usually apply spatial regularization repeatedly throughout the iteration process to smooth DVF, which may result in under-smoothed/oversmoothed DVF.<sup>10,11,15–19</sup> Inaccurate DVF could result in unrealistically deformed images such as falsely deformed ribs or spine.<sup>19</sup>

We have previously developed an unsupervised network for 4D-CT abdominal DIR.<sup>20</sup> In this study, we have extended our previous work to 4D-CT lung DIR. Lung DIR differs from abdominal DIR in (a) lung motion is different from abdominal motion in both motion pattern and motion amplitude. Lung motion are induced by both respiratory motion and cardiac motion. The large motion amplitude of lung diaphragm poses additional challenge to 4D-CT lung DIR; (b) DIR of lung is mainly driven by pulmonary bronchi, fissures, vascular structures, and lung boundaries such as diaphragm. To address these challenges, we proposed a novel deep learning-based DIR method for 4D-CT lung image registration. We trained a network to perform unsupervised direct transformation prediction. The proposed method utilized a new adversarial loss to encourage realistic transformation prediction. In addition, pulmonary vessel enhancement was applied prior to network training to highlight important vasculature structures for accurate lung registration. The major contributions of our work are:

1. Adversarial network was integrated into LungRegNet to enforce additional DVF regularization by penalizing unrealistic deformed image.
2. Pulmonary vessel enhancement prior to network DVF prediction was proposed and proven to be effective.
3. We have achieved the best TRE values so far on DIR-LAB among deep-learning based 4D-CT lung DIR methods.

## 2. RELATED WORKS

Deep learning-based DIR methods have been proposed for MRI brain,<sup>21</sup> CT head/neck,<sup>22</sup> CT chest,<sup>23</sup> MR/US prostate,<sup>24</sup>

4D-CT lung<sup>25–28</sup> and so on.<sup>29</sup> Eppenhof et al. proposed a supervised convolutional neural network (CNN) using U-Net architecture.<sup>27</sup> They trained their network using synthetic random transformations. They have achieved an average TRE value of  $2.17 \pm 1.89$  mm on DIRLAB datasets. One limitation of this study is that the random transformations are very different from the actual lung motion. Supervised training using the random transformations may not provide valid DVF regularization. Another limitation of this study is that the network was trained using whole image, which requires excessive memory. As a result, Eppenhof et al. must perform image down-sampling on the original datasets to save memory, causing image information loss.<sup>27</sup> Sentker et al. developed a general deep learning-based fast image registration called GDL-FIRE.<sup>26</sup> GDL-FIRE was trained in a supervised manner with ‘ground truth’ DVFs generated by three open source DIR frameworks, which are PlastiMatch,<sup>30</sup> NiftyReg<sup>31</sup> and VarReg.<sup>32</sup> To cover large lung motion, GDL-FIRE used iterative DIR by cascading several trained models together. On average, GDL-FIRE was able to achieve TRE of  $2.50 \pm 1.16$  mm on DIRLAB datasets. Besides randomly generated and traditional DIR-generated DVFs, model-based artificial DVF was also used for ground truth transformation generation. Sokooti et al. used model-based respiratory motion to simulate ground truth DVF for their network training.<sup>33</sup> They also generated random transformations with single and mixed frequencies for network training. They trained various network structures including U-Net on whole image and U-Net advanced using image patches. The registration performances were evaluated using TRE and Jacobian determinant. After comparison, they have showed that network trained with model-based respiratory motion outperformed networks trained with random transformations. An average TRE of 2.32 and 1.86 mm were obtained for SPREAD and DIRLAB datasets, respectively. The above-mentioned methods are supervised transformation prediction methods, meaning known ground truth transformations are required for network training. However, it is difficult to artificially generate realistic transformations for network training. To overcome this challenge, unsupervised transformation prediction methods have been proposed. De Vos et al. proposed an unsupervised CNN called deep learning image registration (DLIR).<sup>25</sup> DLIR utilized multistage training and testing. Tested on DIRLAB 4D-CT lung datasets, DLIR produced an average TRE of  $2.64 \pm 4.32$  mm. Loss function of DLIR includes two terms which are NCC between the fixed and deformed images and a bending energy loss term to encourage DVF smoothness. Fechter et al. proposed a one-shot learning for DIR and periodic motion tracking.<sup>34</sup> They employed a UNet with multiresolution approach for coarse-to-fine image registration. The network was trained in an unsupervised manner, meaning no known transformation was required. A spatial transformer network<sup>35</sup> was used to warp the moving image during the training process. The warped images were compared to the fixed image for image dissimilarity loss calculation. Besides the commonly used image dissimilarity loss and transformation smoothness loss, they have



incorporated a cyclic constraint to encourage periodic respiratory motion. They have achieved an average TRE of  $1.83 \pm 2.35$  mm on DIRLAB datasets. Most recently, Jiang et al. proposed an unsupervised transformation prediction method utilizing a multiscale framework for 3D CT lung.<sup>36</sup> Multiple CNN were cascaded to perform multiscale registration with each CNN trained to register on a specific image scale level. The network loss function consists of an image similarity loss and a transformation smoothness loss. The network was trained using image patches. They have demonstrated good network generality by applying the network trained on SPARE datasets to register DIRLAB datasets. In addition, they also showed that the same trained network could be applied on CT-CBCT and CBCT-CBCT registration without retraining or fine-tuning. An average TRE of  $1.66 \pm 1.44$  mm was obtained on DIRLAB datasets.

Generative adversarial networks (GAN) was widely used for image synthesis such as MR to synthetic CT,<sup>37,38</sup> CT to synthetic MR,<sup>39,40</sup> CBCT to synthetic CT,<sup>41</sup> low-dose PET to synthetic full-dose PET,<sup>42</sup> and non-attenuation correction PET to attenuation correction PET.<sup>43</sup> A typical GAN usually consists of a generator and a discriminator. The discriminator was trained to generate an adversarial loss to penalize unrealistic image synthesis. The idea of GAN has been explored in medical image registration as well. The usage of GAN could roughly be categorized into two types, (a) to use adversarial loss for additional transformation regularization and (b) to use GAN to perform image domain translation to cast multimodal image registration to unimodal image registration. Yan et al. used a discriminator to distinguish between predicted transformation-warped image and ground truth transformation-warped image for MR and transrectal ultrasound (TRUS) prostate registration.<sup>44</sup> Ground truth transformation was required and obtained using manually performed registration by experts. The registration was restricted to rigid and affine transformation. Later, Fan et al. proposed an adversarial learning approach for multimodal and unimodal image registration.<sup>45</sup> The discriminator was trained to judge whether an image pair of warped and fixed images was well-aligned. However, paired MR and CT images were required for the discriminator training in multimodal registration. For unimodal registration, they defined a well-aligned warped image as a linearly weighted combination of moving and fixed image which may introduce bias. Hu et al. trained an adversarial network to tell whether a transformation is predicted by the network or generated by finite element method (FEM).<sup>46</sup> The purpose of the adversarial network was to introduce biomechanical constraint to MR and TRUS prostate registration. Another way of utilizing GAN in image registration is to cast multimodal to unimodal image registration via image synthesis. Salehi et al. used a conditional GAN to translate T1 image to T2 image for fetal brain image registration.<sup>47</sup> Qin et al. employed 2D image-to-image translation to disentangle image into domain-invariant latent space prior to registration.<sup>48</sup> A discriminator was trained to distinguish whether an image patch pair was well-aligned or not. Mahapatra et al. combined a conditional GAN and a cyclic GAN for

multimodal image registration.<sup>49</sup> The conditional GAN was used to translate the moving image to the fixed image domain while the cyclic GAN was used to enforce deformation consistency. They tested the network on 2D retina image registration and MR cardiac image registration. Tanner et al. has proposed to use cycle GAN to perform MR-CT image synthesis, followed by traditional DIR with NCC as similarity measures.<sup>50</sup> Manual segmentation was used to evaluate the performance of the registration. The segmentation volume overlap ratio has dropped compared to traditional DIR with normalized mutual information (NMI), modality independent neighborhood descriptor (MIND) due to the poor image synthesis at the thoracic region.

In this study, we have proposed an unsupervised transformation prediction method that is specific to 3D-CT lung DIR. Our method differs from previously mentioned methods in two main aspects, (a) our discriminator was trained to distinguish between the warped image and the fixed image, rather than to distinguish between a ground truth (positive) image alignment and a predicted (negative) image alignment. The advantage is that we could eliminate the requirement of ground truth image alignment by discriminating on images instead of image pairs. For example, ground truth transformations were required in Yan et al.'s work,<sup>44</sup> prealigned MR-CT image pairs were required in Fan et al.'s work,<sup>45</sup> and FEM-generated transformations were required in Hu et al.'s work<sup>46</sup>. In terms of unsupervised learning, recently published Jiang et al.'s work<sup>36</sup> is most relevant to our work, however, they did not use adversarial loss or vessel enhancement. We have achieved better TRE than them on DIRLAB datasets; (b) Pulmonary vasculature enhancement was performed prior to image registration to increase the registration accuracy. NCC was used as image patch similarity measure throughout our study. Inside the lung, vasculature structures are one of the most important driving forces for accurate registration since the intensity of the background lung are small which may induce uncertainty in NCC calculation. Vessel enhancement could increase the relative importance of vessel in NCC calculation, thereby, increasing the robustness of similarity measure. We have demonstrated the efficacy of vessel enhancement in our experiments. The aforementioned deep learning-based methods,<sup>25–27</sup> either supervised or unsupervised, have yet achieved average TRE values of <2 mm on DIRLAB datasets. On the contrary, many conventional intensity-based DIR methods have achieved such accuracy for many years.<sup>17,19,51–53</sup> From this perspective, deep learning-based DIR methods have yet to outperform conventional DIR methods. Deep learning-based methods generally perform very well on classification tasks such as image classification and segmentation. However, deep learning-based registration is a generative task rather than a classification task. To overcome this challenge, we propose to integrate generative adversarial networks (GAN) into LungRegNet. GAN is specialized on generating images instead of classifying images, which fits the task of image registration well.



### 3. MATERIALS AND METHODS

The flowchart of the proposed method was shown in Fig. 1, which consists of three stages which are image preprocessing, training, and inference. Vessel enhancement was first performed as an image preprocessing step. Details of the vessel enhancement is described in Section 3.B. Then, two subnetworks which are CoarseNet and FineNet were trained one by one at the training stage. As the name suggests, CoarseNet predicts large lung motion on a coarse scale image while FineNet predicts local lung motion on a fine scale. Two channel inputs including the fixed and moving image patches were fed into CoarseNet for training. Coarse DVFs were predicted by CoarseNet and used to deform the original moving image patches using a spatial transformer.<sup>35</sup> The warped image was compared to the fixed image for image dissimilarity loss calculation. The CoarseNet and FineNet both include a discriminator which was trained using the warped image and the fixed image to generate adversarial losses for additional regularization. DVF prediction was also regularized by commonly used DVF smoothness constraint loss. At the inference stage, image patches were regularly sampled from the original images by sliding a window of size  $64 \times 64 \times 64$  with overlap size of 8 between neighboring patches. Whole image DVF was obtained by fusing the image patches. Coarse DVF was first predicted using CoarseNet. Then the coarsely warped moving image and fixed image were taken as FineNet input for fine DVF prediction. Final DVF was obtained by summing the coarse DVF and the fine DVF. Detailed network design of the CoarseNet, FineNet, and discriminator was shown in Fig. 3 and described in Section 3.C.

#### 3.A. Datasets

A set of ten 4D-CT lung datasets was retrospectively collected from the authors' department for fivefold cross validation. The original image resolution of the datasets was  $0.97 \times 0.97 \times 2.0$  mm. To compare our results with other DIR methods, public DIRLAB datasets<sup>54,55</sup> were used as separate testing datasets for registration accuracy evaluation. Ten 4D-CT cases with 300 manually selected landmark pairs per case were provided in DIRLAB datasets. The original in-plane image resolutions of DIRLAB datasets range from 0.97–1.16 mm while the slice thickness was 2.5 mm. As a preprocessing step, each dataset was resampled in the superior–inferior direction to match the in-plane resolution. Since we were interested in registration of the lung volumes only, we cropped the images to cover only the lung. To avoid boundary effect, a margin of 24 pixels were preserved after image cropping. Therefore, a total of twenty 4D-CT lung datasets with isotropic resolutions were used in this study.

#### 3.B. Pulmonary vessel enhancement

Computed tomography lung image registration is mainly driven by pulmonary vessels since image intensities of background lung tissue were relatively small compared to that of

pulmonary vessels. Similarly, large vascular branches are usually brighter than small vascular branches in lung. Therefore, large vascular structures could easily outweigh small vascular structures in terms of image similarity loss. Nevertheless, Fig. 2 shows that small vascular structures could spread to the proximity of lung pleural where large vascular structures are absent. Therefore, it could be helpful to increase the weight of small vascular structures in the image similarity loss. Based on this observation, we propose to extract the pulmonary vascular structures to enhance the image contrast of small vascular structures.

To extract vascular structures, the lungs were first automatically segmented in the CT using thresholding and morphological operations. For DIRLAB cases, a threshold value of 700 HU was used to mask the lung. Holes induced by the vasculature structures and bronchi were filled using Matlab built-in functions of 'imfill' and 'imclose' with 3d ball shaped structural element (radius = 10 pixel). After the lung was segmented, a stacked multiscale dictionary learning model (MDLM) was then employed to generate vascular probability maps within the lung.<sup>56</sup> The MDLM method utilized learnt voxel-wise features in a logistic regression classifier to predict the vasculature probability. The code we used for vessel extraction is publicly available at <https://github.com/sepehrehosseiniyah/Segmentation>. The method has won the 1<sup>st</sup> place in VESSEL12 grand challenge. To highlight the vessels to drive lung registration, vessel-enhanced images were generated using the following equation:

$$I_v = I_0 + 1000 * P_v$$

where  $I_v$  represents the vessel-enhanced images,  $I_0$  represents the original images and  $P_v$  is the probability maps of the pulmonary vascular maps.

#### 3.C. Network design

Figure 3 shows the network architecture of the proposed CoarseNet, FineNet, and Discriminator. Both the CoarseNet and FineNet consist of a generator and a discriminator. The generator was trained to directly predict DVF that was used to deform the moving image to match the fixed image. The discriminator was trained to distinguish the deformed images from the original images. The purpose of the discriminator was to impose additional DVF regularization to prevent unrealistic deformed images.

In CoarseNet, image patches were reduced from  $64 \times 64 \times 64$  to  $8 \times 8 \times 8$  through the encoding path using three max pooling layers with kernel sizes of  $2 \times 2 \times 2$  to account for large lung motion. There are totally 12 convolutional layers in the encoding path. The detailed output sizes and feature numbers are shown in Fig. 3. At the beginning of CoarseNet encoding path, three dilated convolutional layers (dilation rates were set to 2, 4, 6, respectively) were used to extract features from the vessel-enhanced input image pairs. The feature maps went through three compression operators to reduce the volume size to 1/8 of original volume size. Each compression operator was implemented by



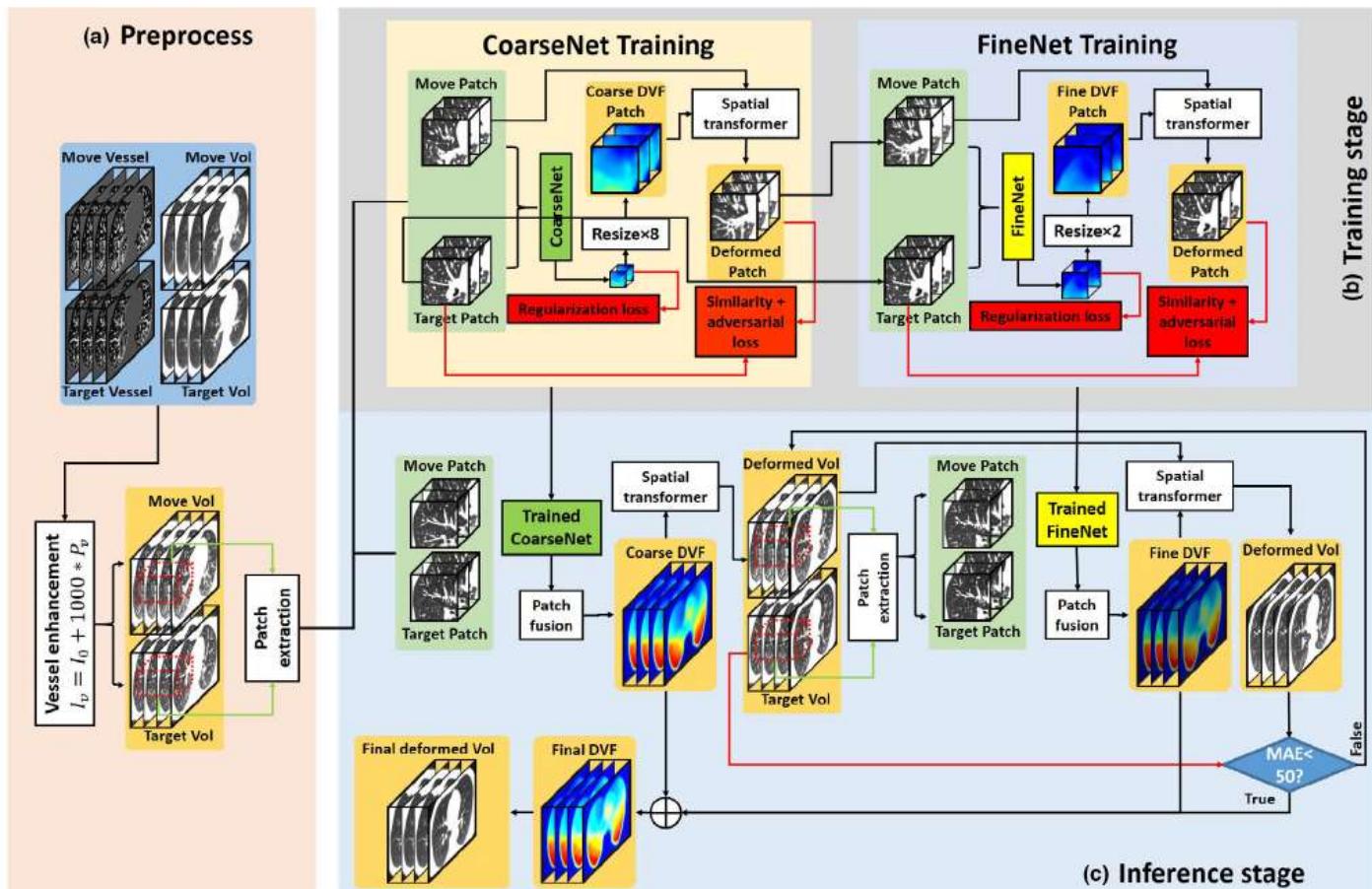


Fig. 1. Workflow of LungRegNet. (a) Preprocessing of pulmonary vessel enhancement. (b) Training stage of the network. (c) Inference stage of the network. [Color figure can be viewed at [wileyonlinelibrary.com](http://wileyonlinelibrary.com)]

several convolution layers with ‘same’ padding and batch normalization and followed by a max-pooling operator. The feature maps were subsequently fed into a convolution layer with batch normalization, followed by a convolution layer for DVF prediction. The number of feature maps were set to 3 to generate coarse DVF prediction in the x, y, and z directions.

The discriminator in CoarseNet consists of four convolutional layers. Each of the convolutional layer was followed by a max-pooling operator to reduce the feature map size. A regression convolution layer with number of feature maps size set to 1 was used in the end to differentiate the deformed image patches from the original image patches. The image patches were first center-cropped to images patches of  $32 \times 32 \times 32$  from  $64 \times 64 \times 64$  before feeding into the discriminator. This is to avoid the boundary effect of registration by ignoring the image patch margins. In addition, two attention gates were used to highlight salient features between two compression operators to focus on the motion between moving and fixed patches.

In FineNet, image patches were reduced from  $32 \times 32 \times 32$  to  $16 \times 16 \times 16$  through the encoding path using six convolutional layers with ‘valid’ padding. Prior to FineNet training, the image patches were center-cropped to  $32 \times 32 \times 32$  from  $64 \times 64 \times 64$  to avoid boundary effect of registration. The final regression convolutional layer was used to generate fine DVF between the coarsely

deformed patches and the fixed patches. Instead of using max pooling layer, several convolution layers with ‘valid’ padding were used to reduce the output DVF size to  $1/2$  as compared to original volume size. Then the predicted DVFs were up-sampled using bicubic interpolation to its original image patch size. As compared the up-sampling rate of 8 in CoarseNet, a smaller up-sampling rate of 2 was used in FineNet to capture relatively smaller lung motion post CoarseNet DVF prediction. The discriminator in FineNet shares a similar architecture with a smaller number of layers as compared to the discriminator in CoarseNet. Similarly, the image patches were first center-cropped to images patches of  $16 \times 16 \times 16$  from  $32 \times 32 \times 32$  before feeding into the discriminator to avoid boundary effect of registration.

### 3.D. Loss function

Loss function of LungRegNet consists of three parts which are the image similarity loss, the adversarial loss, and the regularization loss.

$$G = \min_G \left\{ 1 - NCC(I_{mov}^\circ \varphi, I_{fix}) + \alpha \cdot \text{ADV}(I_{mov}^\circ \varphi, I_{fix}) + \beta \cdot R(\varphi) \right\} \quad (1)$$





FIG. 2. Pulmonary vessel enhancement, a1: three-dimensional rendering of pulmonary vascular probability maps within the lung overlaid on computed tomography image, a2: original axial image, a3: vessel-enhanced a2, a4: original coronal image, a5: vessel-enhanced a4. [Color figure can be viewed at wileyonlinelibrary.com]

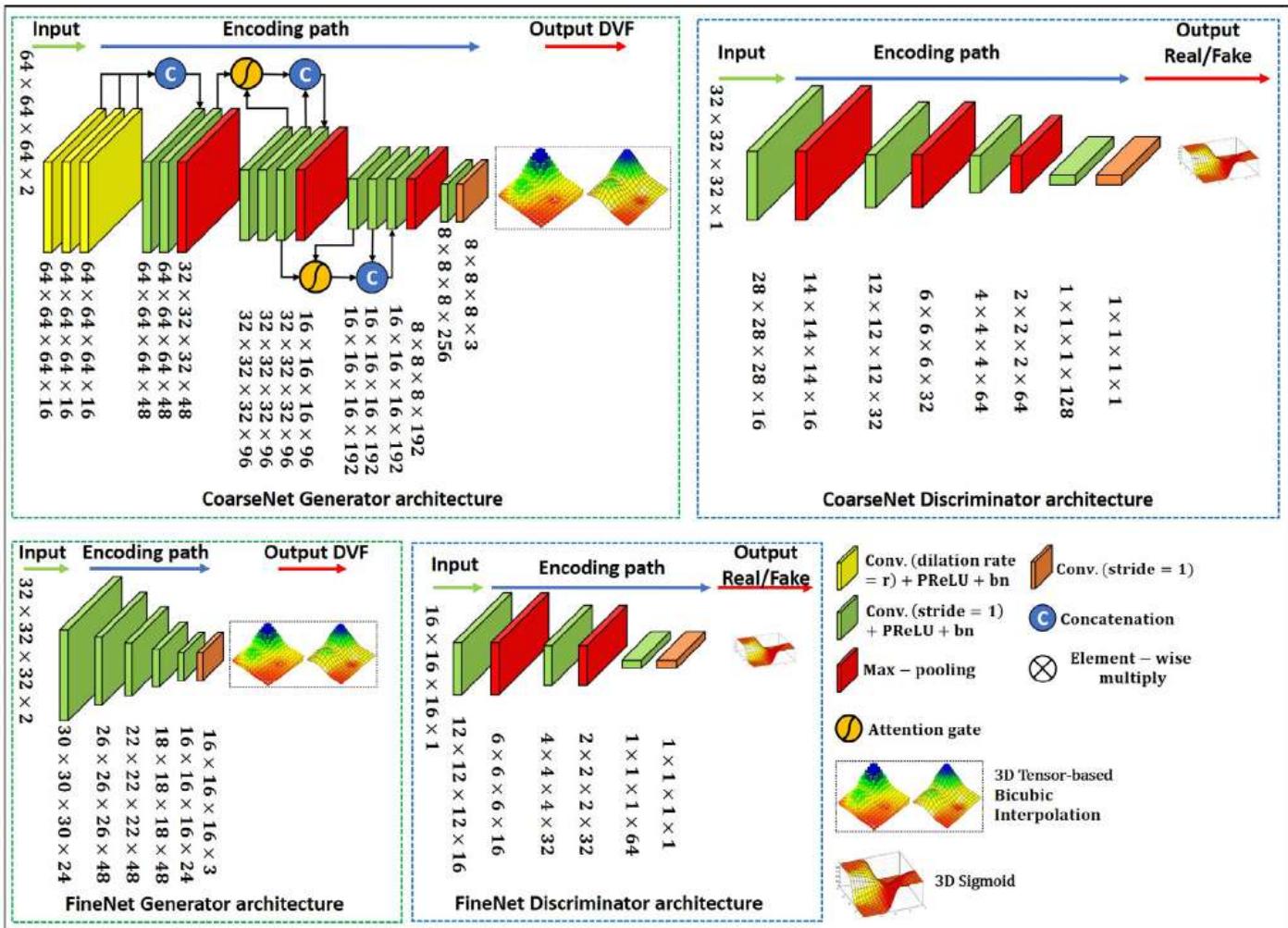


FIG. 3. Detailed network architectures of the CoarseNet, FineNet, and Discriminators. [Color figure can be viewed at wileyonlinelibrary.com]

where  $\varphi = G(I_{mov}, I_{fix})$  represents the predicted DVF for a moving and fixed image pair. The deformed image,  $I_{mov}^\circ, \varphi$ , was obtained by deforming the moving image patch by the predicted deformation field using spatial transformer.  $NCC(\cdot)$  denotes the normalized cross-correlation loss, which is the image similarity loss. To avoid registration boundary effect, we calculated the image similarity loss within only the central region of  $48 \times 48 \times 48$  in the image patches.  $ADV(\cdot)$  denotes the adversarial loss. The adversarial loss was generated by the discriminator and computed as the discriminator binary cross entropy loss of the deformed and fixed images.

Regarding the regularization parameter values, a rule of thumb is that the initial loss terms should be in the same order of magnitude numerically given equal priority. Therefore, we have empirically set the  $\alpha$  and  $\beta$  to be 0.005 and 0.05 respectively.  $R(\varphi)$  denotes the regularization term.

$$R(\varphi) = \mu_1 \|\nabla \varphi\|_2 + \mu_2 \|\nabla^2 \varphi\|_2 \quad (2)$$

The regularization term includes weighted first and second derivatives of the DVF to enforce general smoothness of the predicted DVF. Values of  $\mu_1$  and  $\mu_2$  were empirically set to be 1 and 0.5 in this study.



### 3.E. Training and testing

The LungRegNet was first trained and tested using five-fold cross validation on the ten 4D-CT datasets from our department. Each 4D-CT dataset includes ten phases of 3D-CT throughout a respiratory cycle. In the five-fold cross validation, 4D-CT datasets of eight patients consisting of eighty 3D-CT were taken as training datasets while 4D-CT of the other two patients were used as testing datasets. During training, image pairs between any two phases of the ten phases were taken as the moving and fixed image pairs, which was equivalent to a total of 360 image pairs. The total number of training image pairs were doubled to 720 image pairs after switching the moving and fixed image pairs. The network was trained using patch-based methods by randomly sampling image patches of size  $64 \times 64 \times 64$ .

To compare LungRegNet with other methods on DIRLAB datasets, we trained another network in the same manner except that all the ten 4D-CT datasets from our department were used as training datasets. For testing, phase 50 was registered to phase 00 for each of the ten DIRLAB cases. The reason to train on our datasets and test on separate DIRLAB datasets is to demonstrate the robustness and generalization ability of LungRegNet to datasets from different modality.

Our algorithm was implemented in python 3.6 and TensorFlow on a NVIDIA Tesla V100 GPU with 32 GB of memory. Adam gradient optimizer with learning rate of  $2e-4$  was used for optimization. To deal with large lung motion, we have found out that it is sometimes beneficial to predict fine DVF twice by cascading two trained FineNet in series. Empirically, we chose to perform FineNet twice if the Mean Absolute Error (MAE) value of the Hounsfield Units (HU) between the deformed image and fixed image was greater than 50 within the lung. We have noticed that the registration accuracy barely improved by performing FineNet more than twice.

## 4. RESULTS

### 4.A. Efficacies of discriminator and pulmonary vessel enhancement

To demonstrate the efficacies of the discriminator and vessel enhancement, we compared the registration accuracy for three different variants of LungRegNet, namely LungRegNet-v1 (without discriminator), LungRegNet-v2 (without vessel enhancement), and LungRegNet-v3 (with both discriminator and vessel enhancement). The results shown here are registration between End-Inhalation (EI) as fixed image and End-Exhalation (EE) as moving image. Fused image between the deformed and fixed images were shown in Fig. 4 for all three different variants of LungRegNet. Fig. 4(A1) shows the fused image between fixed and moving images before registration. Fig. 4(A2–A4) show the fused images between fixed and deformed images for LungRegNet-v1, LungRegNet-v2, and LungRegNet-v3 respectively. Figure 4(B1–B4) show similar images as Fig. 4(A1–A4) except that these images were in the

coronal image plane. Fused images were shown as color images with RGB channels represented by the fixed image, the moving image and zero respectively. Yellow color represents equal intensity values of the fixed and deformed images. Arrows in Fig. 4 demonstrated that some misalignments between the fixed and deformed images were observed in LungRegNet-v1 and LungRegNet-v2 but were absent in LungRegNet-v3, indicating improved performance of LungRegNet-v3.

Intensity difference images before and after registration were shown in Fig. 5. We can observe that the intensity differences between the fixed and deformed images were greater for LungRegNet-v1 than that of LungRegNet-v2 and LungRegNet-v3, indicating that the integration of discriminator was very effective in improving the overall lung alignment. To further evaluate the usefulness of pulmonary vessel enhancement, we calculated the mean absolute error (MAE) between the fixed and deformed images after image registration. The MAE results are shown in Table I. We can observe that the MAEs were the smallest for LungRegNet-v3. A comparison between LungRegNet-v1 and LungRegNet-v2 shows that the discriminator was more effective than the vessel enhancement in increasing the overall registration accuracy.

To calculate the target registration error (TRE), we manually selected 20 landmark pairs between the EI and EE phases for each of our ten 4D-CT datasets. The 20 landmark pairs were placed on multiple slices, ranging from the diaphragm to the apex of lung. Landmarks were carefully placed at the bifurcations of pulmonary vessels with good image contrast. Target registration error values were calculated and reported in Table II. On average, TRE values were  $2.59 \pm 1.55$ ,  $1.35 \pm 1.15$ , and  $1.00 \pm 0.53$  for LungRegNet-v1, LungRegNet-v2, and LungRegNet-v3 respectively. The results shown in Table II were in good agreement with Table I.

### 4.B. Registration accuracies on DIRLAB datasets

To compare with other DIR methods, we performed DIR on 10 DIRLAB datasets that are publicly available at (<https://www.dir-lab.com>). The accuracy was evaluated using 300 landmark pairs per case that were provided in the DIRLAB datasets. Quantitative results of deep learning-based DIRs and traditional DIRs are reported in Tables III and IV respectively. On average, the TRE for LungRegNet-v3 was  $1.59 \pm 1.58$  mm, which is much more accurate than the other six deep learning-based methods. Compared with the nine conventional methods shown in Table IV, only Starling et al.<sup>52</sup> and Heinrich et al.<sup>51</sup> outperformed LungRegNet-v3. Given its unsupervised and noniterative nature, LungRegNet-v3 is considered to have excellent performance in terms of TRE. We performed twice DVF predictions on FineNet for cases 6–8 as the MAE between fixed and deformed images after the first FineNet was  $>50$  within the lung. Cases 6–8 are usually more difficult to register than other cases since the TRE values of cases 6–8 before registration was greater than other cases, indicating larger lung motion.



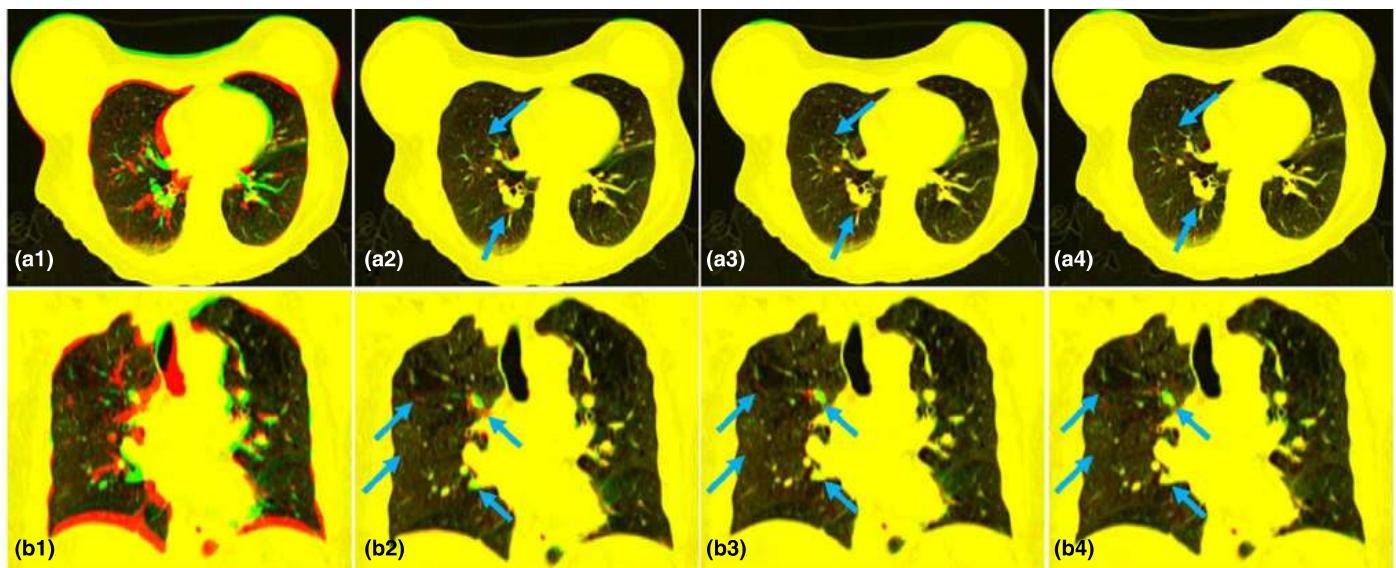


FIG. 4. Fused images, (a1, b1): between fixed and moving images before registration, (a2–a4, b2–b4) between fixed and deformed images for LungRegNet-v1, LungRegNet-v2, and LungRegNet-v3 respectively. Red color represents fixed image while green color represents moving image. Yellow color means intensity agreement, indicating good alignment between fixed and deformed images. [Color figure can be viewed at wileyonlinelibrary.com]

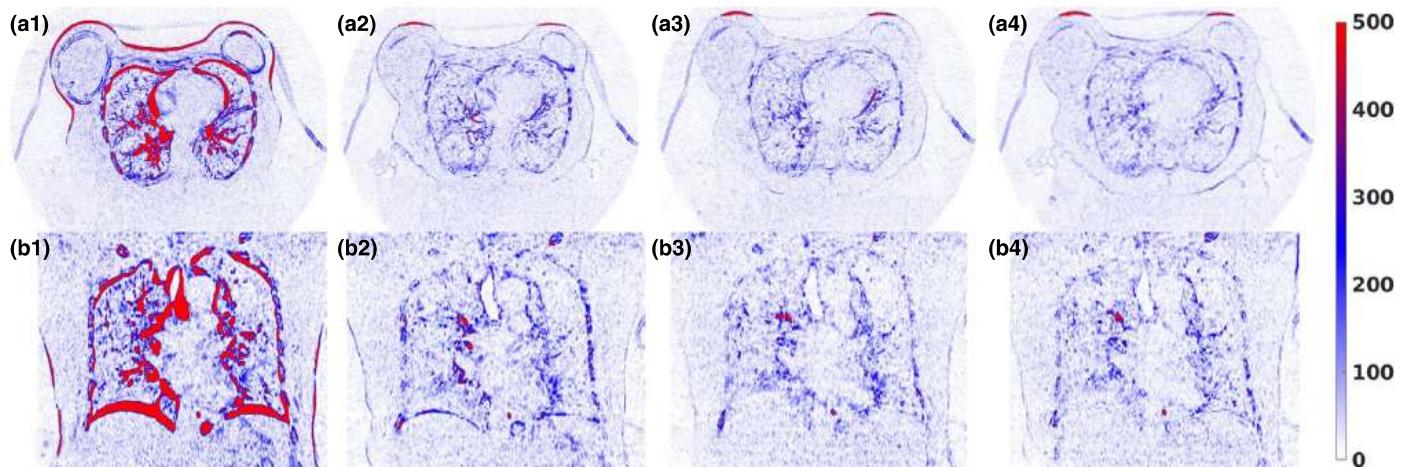


FIG. 5. Intensity difference images, (a1, b1): between the fixed and moving images before registration, (a2–a4, b2–b4) between the fixed and deformed images for LungRegNet-v1, LungRegNet-v2, and LungRegNet-v3 respectively. [Color figure can be viewed at wileyonlinelibrary.com]

TABLE I. Mean absolute error between the fixed image and deformed image for three different variants of LungRegNet using our datasets.

Our datasets	1	2	3	4	5	6	7	8	9	10	Mean
LungRegNet-v1	62.9	114.9	64.1	114.3	59.6	92.5	54.4	47.4	65.6	49.1	72.5 ± 25.4
LungRegNet-v2	54.3	102.7	49.6	91.8	49.5	78.0	42.9	40.0	51.8	41.8	60.2 ± 22.4
LungRegNet-v3	50.2	87.6	41.8	77.6	43.4	65.2	41.4	33.9	44.6	35.5	52.1 ± 18.4

For comparison, DIR using a multiple-stage conventional Horn-Schunck optical flow method with Gaussian smoothing filter was performed on DIRLAB datasets. The optical flow method used five stages with 20 passes in each stage and 20 iterations in each pass, which is equivalent to a total of 2000 iterations. Figure 6 illustrates the results on the most challenge case 8 of DDIRLAB. DVF magnitude colormaps were overlaid on top of CT images. The arrows indicate that lung motion field failed to propagate to some regions due to either

the lack of image contrast or the existence of local lesions. On the contrary, the DVF of LungRegNet was smooth and realistic.

## 5. DISCUSSIONS

Conventional DIR methods usually require careful parameter tuning to maximize its performance, which depends on the image similarity metrics used, registration anatomical



sites and imaging modalities<sup>6</sup>. It often takes multiple experiments and evaluations before a set of optimal parameters are found. The iterative nature of conventional DIRs makes them computationally demanding especially for large image datasets since spatial filtering of DVF was often applied repeatedly during the optimization process. To address these issues, we proposed a novel deep learning-based DIR method for fast, robust and accurate DIR of 4D-CT lung.

LungRegNet was trained using unsupervised and patch-based method. One benefit of unsupervised training is that it could mitigate the problems of lack of ground truth DVFs and the shortage of training datasets. Since ground truth DVF is not required, any 4D-CT datasets could be used as training dataset. The patch-based training could also alleviate the problem of shortage of training datasets since a large set of image patches could be sampled from the original 4D-CT. Another advantage of patch-based method is that it allows the original image to be up-sampled prior to DVF prediction

TABLE II. Comparison of target registration error (TRE) values for three different variants of LungRegNet.

Our datasets	TRE before registration	TRE after registration		
		LungRegNet-v1	LungRegNet-v2	LungRegNet-v3
1	7.91 ± 3.01	1.17 ± 0.83	1.13 ± 1.00	0.89 ± 0.59
2	10.70 ± 4.06	5.15 ± 3.75	2.36 ± 3.46	2.11 ± 1.73
3	8.77 ± 4.44	1.29 ± 0.79	0.71 ± 0.26	0.76 ± 0.54
4	14.85 ± 3.19	5.41 ± 4.36	2.41 ± 1.31	1.59 ± 0.37
5	7.89 ± 4.21	1.45 ± 0.85	0.60 ± 0.50	0.46 ± 0.23
6	14.84 ± 1.39	7.00 ± 2.58	2.88 ± 3.25	1.37 ± 0.25
7	5.89 ± 3.39	1.14 ± 0.20	0.89 ± 0.12	0.66 ± 0.23
8	11.12 ± 1.98	0.92 ± 0.47	0.74 ± 0.18	0.69 ± 0.49
9	6.97 ± 4.02	1.11 ± 0.71	0.85 ± 0.53	0.81 ± 0.40
10	6.96 ± 1.31	1.29 ± 0.99	0.96 ± 0.87	0.64 ± 0.42
Mean	9.59 ± 3.10	2.59 ± 1.55	1.35 ± 1.15	1.00 ± 0.53

since patch-based training, as opposed to whole-image based training, was not GPU memory intensive. Image up-sampling prior to DVF prediction could help increase registration accuracy. Our excellent TRE values could be partially due to the fact that images were up-sampled in the superior-inferior direction to match its in-plane resolution prior to DVF prediction. To date, our TRE values on DIRLAB datasets are the most accurate among the published deep learning-based DIR methods on 4D-CT. It is worth noting that our excellent TRE values on DIRLAB were achieved using network that was trained using non-DIRLAB datasets. This is a strong evidence that LungRegNet has great ability to generalize well on datasets from different modalities. To achieve great robustness, it is important to normalize the input images to zero mean and unit variance prior to network training and testing.

Since the CoarseNet was designed to predict coarse DVF, a down-sampling rate of 8 was implemented by using three max pooling layers prior to convolutional layers. Given that normal diaphragmatic motion ranges 3–5 cm in the superior-inferior direction and image slice thickness ~1 mm (with image resampling), the maximum diaphragmatic motion could be represented by DVF prediction with maximum value of 6.25 with down-sampling rate of 8. The network takes image patches with matrix size of 64 × 64 × 64 as input. Image patches were resized to 8 × 8 × 8 after these max pooling layers. To generate DVFs with consistent matrix sizes as the input image patches, one option was to stack three transpose-convolutional layers to resize the DVF from 8 × 8 × 8 to 64 × 64 × 64. Another option is to use bicubic interpolation to resize the DVF. Through our experiments, we have found out that bicubic interpolation with no trainable parameters performed much better than the transpose-convolutional layers in predicting realistic DVF. This is because bicubic interpolation could generate smooth DVFs that model the actual lung motion more accurately than transpose-convolutional layers.

Different from many supervised deep learning methods, we trained the LungRegNet in a completely unsupervised

TABLE III. Target registration error values for different deep learning-based methods on DIRLAB datasets, bold values demonstrate cases where FineNet was performed twice.

Set	Before registration	Deep learning-based methods						
		Eppenhof et al. <sup>27</sup>	De Vos et al. <sup>25</sup>	Sentker et al. <sup>26</sup>	Fechter et al. <sup>34</sup>	Sokooti et al. <sup>33</sup>	Jiang et al. <sup>36</sup>	LungRegNet-v3
1	3.89 ± 2.78	1.45 ± 1.06	1.27 ± 1.16	1.20 ± 0.60	1.21 ± 0.88	1.13 ± 0.51	1.20 ± 0.63	0.98 ± 0.54
2	4.34 ± 3.90	1.46 ± 0.76	1.20 ± 1.12	1.19 ± 0.63	1.13 ± 0.65	1.08 ± 0.55	1.13 ± 0.56	0.98 ± 0.52
3	6.94 ± 4.05	1.57 ± 1.10	1.48 ± 1.26	1.67 ± 0.90	1.32 ± 0.82	1.33 ± 0.73	1.30 ± 0.70	1.14 ± 0.64
4	9.83 ± 4.86	1.95 ± 1.32	2.09 ± 1.93	2.53 ± 2.01	1.84 ± 1.76	1.57 ± 0.99	1.55 ± 0.96	1.39 ± 0.99
5	7.48 ± 5.51	2.07 ± 1.59	1.95 ± 2.10	2.06 ± 1.56	1.80 ± 1.60	1.62 ± 1.30	1.72 ± 1.28	1.43 ± 1.31
6	10.89 ± 6.9	3.04 ± 2.73	5.16 ± 7.09	2.90 ± 1.70	2.30 ± 3.78	2.75 ± 2.91	2.02 ± 1.70	<b>2.26 ± 2.93</b>
7	11.03 ± 7.4	3.41 ± 2.75	3.05 ± 3.01	3.60 ± 2.99	1.91 ± 1.65	2.34 ± 2.32	1.70 ± 1.03	<b>1.42 ± 1.16</b>
8	15.0 ± 9.01	2.80 ± 2.46	6.48 ± 5.37	5.29 ± 5.52	3.47 ± 5.00	3.29 ± 4.32	2.64 ± 2.78	<b>3.13 ± 3.77</b>
9	7.92 ± 3.98	2.18 ± 1.24	2.10 ± 1.66	2.38 ± 1.46	1.47 ± 0.85	1.86 ± 1.47	1.51 ± 0.94	1.27 ± 0.94
10	7.3 ± 6.35	1.83 ± 1.36	2.09 ± 2.24	2.13 ± 1.88	1.79 ± 2.24	1.63 ± 1.29	1.79 ± 1.61	1.93 ± 3.06
Mean	8.46 ± 5.48	2.17 ± 1.89	2.64 ± 4.32	2.50 ± 1.16	1.83 ± 2.35	1.86 ± 2.12	1.66 ± 1.44	1.59 ± 1.58



TABLE IV. Target registration error values for the-state-of-art traditional methods on DIRLAB datasets.

Set	Before registration	Conventional methods								MIND Li et al. <sup>59</sup>	ANTs+NCC Heinrich et al. <sup>60</sup>
		Schmidt-Richberg et al. <sup>57</sup>	Heinrich et al. <sup>51</sup>	Fu et al. <sup>19</sup>	Delmon et al. <sup>17</sup>	Vandemeulebroucke et al. <sup>16</sup>	Staring et al. <sup>52</sup>	Elastix+MI Gong et al. <sup>58</sup>			
1.	3.89 ± 2.78	1.22 ± 0.64	0.97 ± 0.5	1.06 ± 0.50	1.2 ± 0.6	1.52 ± 0.92	0.99 ± 0.57	1.23	1.05	—	
2.	4.34 ± 3.90	1.14 ± 0.65	0.96 ± 0.5	1.09 ± 0.57	1.1 ± 0.6	1.30 ± 1.03	0.94 ± 0.53	1.15	1.06	—	
3.	6.94 ± 4.05	1.36 ± 0.81	1.21 ± 0.7	1.51 ± 1.00	1.6 ± 0.9	1.69 ± 1.12	1.13 ± 0.64	1.92	1.23	—	
4.	9.83 ± 4.86	2.68 ± 2.79	1.39 ± 1.0	1.73 ± 1.55	1.6 ± 1.1	1.82 ± 1.14	1.49 ± 1.01	1.81	1.48	—	
5.	7.48 ± 5.51	1.57 ± 1.23	1.72 ± 1.6	1.80 ± 1.63	2.0 ± 1.6	2.75 ± 2.45	1.77 ± 1.53	2.52	1.62	—	
6.	10.89 ± 6.9	2.21 ± 1.66	1.49 ± 1.0	2.25 ± 2.61	1.7 ± 1.0	2.01 ± 1.16	1.29 ± 0.85	2.42	1.61	—	
7.	11.03 ± 7.4	3.81 ± 3.06	1.58 ± 1.2	1.41 ± 0.98	1.9 ± 1.2	2.15 ± 1.59	1.26 ± 1.09	3.92	2.04	—	
8.	15.0 ± 9.01	3.42 ± 4.25	2.11 ± 2.4	3.53 ± 5.70	2.2 ± 2.3	2.11 ± 1.79	1.87 ± 2.57	5.65	3.46	—	
9.	7.92 ± 3.98	1.83 ± 1.19	1.36 ± 0.7	2.31 ± 1.88	1.6 ± 0.9	2.05 ± 1.20	1.33 ± 0.98	3.05	1.37	—	
10.	7.3 ± 6.35	2.06 ± 1.92	1.43 ± 1.6	1.18 ± 1.97	1.7 ± 1.2	2.12 ± 1.66	1.14 ± 0.89	2.29	1.63	—	
Mean	8.46 ± 5.48	2.13 ± 1.82	1.43 ± 1.3	1.78 ± 1.83	1.66 ± 1.14	1.95 ± 1.47	1.32 ± 1.24	2.60 ± 1.35	1.66	2.43 ± 4.1	

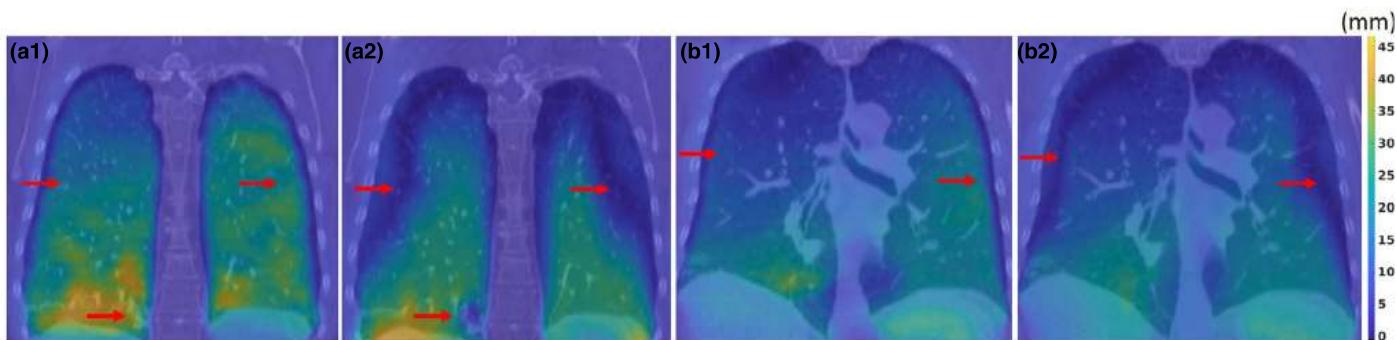


FIG. 6. Deformation vector field (DVF) magnitude colormaps overlaid on computed tomography images for case 8. (a1, b1) show the results using LungRegNet, (a2, b2) show the results using conventional multiple stage Horn-Schunck optical flow method. The arrows indicate regions where DVF from LungRegNet was smoother and more realistic than the optical flow method. [Color figure can be viewed at wileyonlinelibrary.com]

manner. One benefit of unsupervised training is that it could mitigate the problem of unavailability of ground truth. However, unsupervised training faces the challenge of lack of DVF regularization. To overcome this challenge, we proposed to use both discriminator and smoothness constraints for DVF regularization. The discriminator was trained to differentiate the deformed images from the fixed images, generating a loss term by penalizing unrealistic deformed images. Therefore, LungRegNet was encouraged to predict realistic DVFs. It is noteworthy that the discriminator will not affect the inference speed as it was used only in the training stage.

One limitation of this study is that sliding motion at the lung pleura was not supported well, evidenced by small DVF values at the lung pleura (Fig. 6). This is because (a) the ribs near the pleura barely move, causing DVF to be small, (b) there are less pulmonary vascular structures near the lung pleura, leading to minimal driving force for the lung motion. Conventional DIR methods model the sliding motion by applying direction-dependent spatial filters repeatedly.<sup>17,19,57</sup> Since LungRegNet is a noniterative method, we plan to integrate biomechanical model into DVF regularization to model the sliding motion in the future.<sup>61</sup>

## 6. CONCLUSIONS

An unsupervised deep learning-based method was developed for 4D-CT lung DIR. The proposed method was able to accurately register images between any two 4D-CT phases within 1 min using powerful NVIDIA Tesla V100 GPU. The proposed LungRegNet is a promising tool for lung motion management and treatment planning during radiation therapy.

## ACKNOWLEDGMENT

This research is supported in part by the National Cancer Institute of the National Institutes of Health under Award Number R01CA215718, and Dunwoody Golf Club Prostate Cancer Research Award, a philanthropic award provided by the Winship Cancer Institute of Emory University.

## CONFLICT OF INTEREST

Dr. Kristin Higgins is consulting for Astra Zeneca, Varian, on advisory board for Genentech, and receiving research funding from RefleXion Medical.



\*Co-first author.

<sup>a</sup>Author to whom correspondence should be addressed. Electronic mail: xiaofeng.yang@emory.edu; Telephone: (404)-778-8622; Fax: (404)-778-4139.

## REFERENCES

- Yang D, Lu W, Low DA, Deasy JO, Hope AJ, El Naqa I. 4D-CT motion estimation using deformable image registration and 5D respiratory motion modeling. *Med Phys*. 2008;35:4577–4590.
- Li F, Li J, Ma Z, et al. Comparison of internal target volumes defined on 3-dimensional, 4-dimensional, and cone-beam CT images of non-small-cell lung cancer. *Oncotarget Ther*. 2016;9:6945–6951.
- Brock KK, Mutic S, McNutt TR, Li HH, Kessler ML. Use of image registration and fusion algorithms and techniques in radiotherapy: report of the AAPM Radiation Therapy Committee Task Group No. 132. *Med Phys*. 2017;44:e43–e76.
- Lin H, Lu H, Shu L, et al. Dosimetric study of a respiratory gating technique based on four-dimensional computed tomography in non-small-cell lung cancer. *J Radiat Res*. 2014;55:583–588.
- MacManus M, Everitt S, Schimek-Jasch T, Li XA, Nestle U, Kong F-MS. Anatomic, functional and molecular imaging in lung cancer precision radiation therapy: treatment response assessment and radiation therapy personalization. *Transl Lung Cancer Res*. 2017;6:670–688.
- Sotiras A, Davatzikos C, Paragios N. Deformable medical image registration: a survey. *IEEE Trans Med Imaging*. 2013;32:1153–1190.
- [7]Haskins G, Kruger U, Yan P. Deep Learning in Medical Image Registration: A Survey. ArXiv; abs/1903.02026.2019.
- Shi C, Tang X, Chan M. Evaluation of the new respiratory gating system. *Precis Radiat Oncol*. 2017;1:127–133.
- Yang D, Li H, Low DA, Deasy JO, Naqa IEI. A fast inverse consistent deformable image registration method based on symmetric optical flow computation. *Phys Med Biol*. 2008;53:6143–6165.
- Yang D, Brame S, El Naqa I, et al. Technical note: DIRART—A software suite for deformable image registration and adaptive radiotherapy research. *Med Phys*. 2011;38:67–77.
- Yang D, Goddu SM, Lu W, et al. Technical note: deformable image registration on partially matched images for radiotherapy applications. *Med Phys*. 2010;37:141–145.
- Cahill ND, Noble JA, Hawkes DJ. A Demons algorithm for image registration with locally adaptive regularization. *Med Image Comput Comput Assist Interv*. 2009;12:574–581.
- Vercauteren T, Pennec X, Perchant A, Ayache N. Diffeomorphic demons: Efficient non-parametric image registration. *NeuroImage*. 2009;45:s61–s72.
- Gorbunova V, Sporrin J, Lo P, et al. Mass preserving image registration for lung CT. *Med Image Anal*. 2012;16:786–795.
- Ruan D, Esedoglu S, Fessler JA. Discriminative sliding preserving regularization in medical image registration. Paper presented at: 2009 IEEE International Symposium on Biomedical Imaging: From Nano to Macro; 28 June–1 July 2009, 2009.
- Vandemeulebroucke J, Bernard O, Rit S, Kybic J, Clarysse P, Sarrut D. Automated segmentation of a motion mask to preserve sliding motion in deformable registration of thoracic CT. *Med Phys*. 2012;39:1006–1015.
- Delmon V, Rit S, Pinho R, Sarrut D. Registration of sliding objects using direction dependent B-splines decomposition. *Phys Med Biol*. 2013;58:1303–1314.
- Pace DF, Aylward SR, Niethammer M. A locally adaptive regularization based on anisotropic diffusion for deformable image registration of sliding organs. *IEEE Trans Med Imaging*. 2013;32:2114–2126.
- Fu Y, Liu S, Li HH, Li H, Yang D. An adaptive motion regularization technique to support sliding motion in deformable image registration. *Med Phys*. 2018;45:735–747.
- Lei Y, Fu Y, Harms J, et al. 4D-CT Deformable Image Registration Using an Unsupervised Deep Convolutional Neural Network. Paper presented at: AIRT@MICCAI2019.
- Wu G, Kim M, Wang Q, Munsell BC, Shen D. Scalable high-performance image registration framework by unsupervised deep feature representations learning. *IEEE Trans Biomed Eng*. 2016;63:1505–1516.
- Kearney V, Haaf S, Sudhyadhom A, Valdes G, Solberg TD. An unsupervised convolutional neural network-based algorithm for deformable image registration. *Phys Med Biol*. 2018;63:185017.
- Jiang P, Shackleford JA. CNN Driven Sparse Multi-level B-Spline Image Registration. 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition; 2018:9281–9289.
- Hu Y, Modat M, Gibson E, et al. Weakly-supervised convolutional neural networks for multimodal image registration. *Med Image Anal*. 2018;49:1–13.
- Vos BDd, Berendsen FF, Viergever MA, Staring M, Isgum I. End-to-End Unsupervised Deformable Image Registration with a Convolutional Neural Network. Paper presented at: DLMIA/ML-CDS@MICCAI2017.
- Sentker T, Madesta F, Werner R. GDL-FIRE ^text 4D : Deep Learning-Based Fast 4D CT Image Registration. Paper presented at: MICCAI; 2018.
- Eppenhof KAJ, Pluim JPW. Pulmonary CT registration through supervised learning with convolutional neural networks. *IEEE Trans Med Imaging*. 2019;38:1097–1105.
- Fu Y, Wu X, Thomas AM, Li HH, Yang D. Automatic large quantity landmark pairs detection in 4DCT lung images. *Med Phys*. 2019;46:4490–4501.
- Fu Y, Lei Y, Wang T, Curran WJ, Liu T, Yang X. Deep Learning in Medical Image Registration: A Review. ArXiv; 2019;abs/1912.12318.
- Modat M, Ridgway GR, Taylor ZA, et al. Fast free-form deformation using graphics processing units. *Comput Methods Programs Biomed*. 2010;98:278–284.
- Shackleford JA, Kandasamy N, Sharp GC. On developing B-spline registration algorithms for multi-core processors. *Phys Med Biol*. 2010;55:6329–6351.
- Werner R, Schmidt-Richberg A, Handels H, Ehrhardt J. Estimation of lung motion fields in 4D CT data by variational non-linear intensity-based registration: a comparison and evaluation study. *Phys Med Biol*. 2014;59:4247–4260.
- Sokooti H, Vos BDd, Berendsen FF, et al. 3D Convolutional Neural Networks Image Registration Based on Efficient Supervised Learning from Artificial Deformations. ArXiv. 2019;abs/1908.10235.
- Fechter T, Baltas D. One Shot Learning for Deformable Medical Image Registration and Periodic Motion Tracking. ArXiv. 2019;abs/1907.04641.
- Jaderberg M, Simonyan K, Zisserman A, Kavukcuoglu K. Spatial Transformer Networks. ArXiv. 2015;abs/1506.02025.
- Jiang Z, Yin FF, Ge Y, Ren L. A multi-scale framework with unsupervised joint training of convolutional neural networks for pulmonary deformable image registration. *Phys Med Biol*. 2019;65:015011.
- Lei Y, Harms J, Wang T, et al. MRI-based synthetic CT generation using semantic random forest with iterative refinement. *Phys Med Biol*. 2019;64:085001.
- Lei Y, Harms J, Wang T, et al. MRI-only based synthetic CT generation using dense cycle consistent generative adversarial networks. *Med Phys*. 2019;46:3565–3581.
- Lei Y, Dong X, Tian Z, et al. CT prostate segmentation based on synthetic MRI-aided deep attention fully convolution network. *Med Phys*. 2019;47:530–540.
- Dong X, Lei Y, Tian S, et al. Synthetic MRI-aided multi-organ segmentation on male pelvic CT using cycle consistent deep attention network. *Radiother Oncol*. 2019;141:192–199.
- Harms J, Lei Y, Wang T, et al. Paired cycle-GAN-based image correction for quantitative cone-beam computed tomography. *Med Phys*. 2019;46:3998–4009.
- Lei Y, Dong X, Wang T, et al. Whole-body PET estimation from low count statistics using cycle-consistent generative adversarial networks. *Phys Med Biol*. 2019;64:215017.
- Dong X, Lei Y, Wang T, et al. Deep learning-based attenuation correction in the absence of structural information for whole-body PET imaging. *Phys Med Biol*. 2019. [Epub ahead of print]. <https://doi.org/10.1088/1361-6560/ab652c>
- Yan P, Xu S, Rastinehad AR, Wood BJ. Adversarial Image Registration with Application for MR and TRUS Image Fusion; 2018, Cham.
- Fan J, Cao X, Wang Q, Yap P-T, Shen D. Adversarial learning for mono- or multi-modal registration. *Med Image Anal*. 2019;58:101545.
- Hu Y, Gibson E, Ghavami N, et al. Adversarial Deformation Regularization for Training Image Registration Neural Networks. Paper presented at: MICCAI; 2018.
- Salehi SSM, Khan S, Erdogmus D, Gholipour A. Real-time Deep Registration With Geodesic Loss. ArXiv. 2018;abs/1803.05982.



48. Qin C, Shi B, Liao R, Mansi T, Rueckert D, Kamen A. Unsupervised Deformable Registration for Multi-modal Images via Disentangled Representations; 2019; Cham.
49. Mahapatra D, Sedai S, Garnavi R. Elastic Registration of Medical Images With GANs. ArXiv. 2018;abs/1805.02369.
50. Tanner C, Özdemir F, Profanter R, Vishnevsky V, Konukoglu E, Göksel O. Generative Adversarial Networks for MR-CT Deformable Image Registration. ArXiv. 2018;abs/1807.07349.
51. Heinrich MP, Jenkinson M, Brady M, Schnabel JA. MRF-based deformable registration and ventilation estimation of lung CT. *IEEE Trans Med Imaging*. 2013;32:1239–1248.
52. Staring M, Klein S, Niessen WJ, Stoel BC, Mc E. Pulmonary Image Registration with elastix using a Standard Intensity-Based Algorithm; 2010.
53. Vishnevskiy V, Gass T, Szekely G, Tanner C, Göksel O. Isotropic total variation regularization of displacements in parametric image registration. *IEEE Trans Med Imaging*. 2017;36:385–395.
54. Castillo E, Castillo R, Martinez J, Shenoy M, Guerrero T. Four-dimensional deformable image registration using trajectory modeling. *Phys Med Biol*. 2010;55:305–327.
55. Castillo R, Castillo E, Guerra R, et al. A framework for evaluation of deformable image registration spatial accuracy using large landmark point sets. *Phys Med Biol*. 2009;54:1849–1870.
56. Kiros R, Popuri K, Cobzas D, Jagersand M. Stacked Multiscale Feature Learning for Domain Independent Medical Image Segmentation. Paper presented at: Machine Learning in Medical Imaging; 2014; Cham.
57. Schmidt-Richberg A, Werner R, Handels H, Ehrhardt J. Estimation of slipping organ motion by registration with direction-dependent regularization. *Med Image Anal*. 2012;16:150–159.
58. Gong L, Zhang C, Duan L, et al. Nonrigid image registration using spatially region-weighted correlation ratio and GPU-acceleration. *IEEE J Biomed Health Inform*. 2018;23:766–778.
59. Li Z, van Vliet LJ, Vos FM. *Similarity Image Registration Based on Reorientation of the Hessian*; 2013; Berlin, Heidelberg.
60. Heinrich MP. Deformable lung registration for pulmonary image analysis of MRI and CT scans; 2013.
61. Han L, Dong H, McClelland JR, Han L, Hawkes DJ, Barratt DC. A hybrid patient-specific biomechanical model based image registration method for the motion estimation of lungs. *Med Image Anal*. 2017;39:87–100.

