



An unsupervised image registration method employing chest computed tomography images and deep neural networks

Thao Thi Ho ^a, Woo Jin Kim ^b, Chang Hyun Lee ^{c,d}, Gong Yong Jin ^e, Kum Ju Chae ^e, Sanghun Choi ^{a,*}

^a School of Mechanical Engineering, Kyungpook National University, Daegu, South Korea

^b Department of Internal Medicine and Environmental Health Center, Kangwon National University Hospital, School of Medicine, Kangwon National University, Chuncheon, South Korea

^c Department of Radiology, Seoul National University, College of Medicine, Seoul National University Hospital, Seoul, South Korea

^d Department of Radiology, College of Medicine, The University of Iowa, Iowa City, IA, USA

^e Department of Radiology, Research Institute of Clinical Medicine of Jeonbuk National University, Biomedical Research Institute of Jeonbuk National University Hospital, Jeonju, South Korea

ARTICLE INFO

Keywords:

Deep learning
Unsupervised learning
Image registration
CT lung

ABSTRACT

Background: Deformable image registration is crucial for multiple radiation therapy applications. Fast registration of computed tomography (CT) lung images is challenging because of the large and nonlinear deformation between inspiration and expiration. With advancements in deep learning techniques, learning-based registration methods are considered efficient alternatives to traditional methods in terms of accuracy and computational cost.

Method: In this study, an unsupervised lung registration network (LRN) with cycle-consistent training is proposed to align two acquired CT-derived lung datasets during breath-holds at inspiratory and expiratory levels without utilizing any ground-truth registration results. Generally, the LRN model uses three loss functions: image similarity, regularization, and Jacobian determinant. Here, LRN was trained on the CT datasets of 705 subjects and tested using 10 pairs of public CT DIR-Lab datasets. Furthermore, to evaluate the effectiveness of the registration technique, target registration errors (TREs) of the LRN model were compared with those of the conventional algorithm (sum of squared tissue volume difference; SSTVD) and a state-of-the-art unsupervised registration method (VoxellMorph).

Results: The results showed that the LRN with an average TRE of 1.78 ± 1.56 mm outperformed VoxellMorph with an average TRE of 2.43 ± 2.43 mm, which is comparable to that of SSTVD with an average TRE of 1.66 ± 1.49 mm. In addition, estimating the displacement vector field without any folding voxel consumed less than 2 s, demonstrating the superiority of the learning-based method with respect to fiducial marker tracking and the overall soft tissue alignment with a nearly real-time speed.

Conclusions: Therefore, this proposed method shows significant potential for use in time-sensitive pulmonary studies, such as lung motion tracking and image-guided surgery.

1. Introduction

Image registration is a process, by which optimal spatial mapping is established, bridging moving and target images collected from different perspectives at different times using different imaging modalities or even different subjects [1]. Moreover, lung deformation image registration (DIR) has been widely used in numerous medical applications [1], such as lung segmentation [2,3], lung atlas establishment, target definition, lung tissue motion tracking [4], lung fusion [5],

radiotherapy, and guided lung surgeries [6,7]. In general, traditional image registration has been used to identify a spatial correspondence between images, which iteratively updates the deformation field for an image pair based on a combination of transformation model, cost function, and optimization [8–11]. Here, each iteration has been registered as a progressive optimization problem. These algorithms achieve high accuracy and diffeomorphic properties; however, they are time-consuming, not modality-specific to optimization problems, and provide highly non-convex optimization results. Hence, lung registration via conventional image registration is a challenging task because of

* Corresponding author. School of Mechanical Engineering, Kyungpook National University, 80 Daehak-ro, Buk-gu, Daegu, 41566, South Korea.

E-mail address: s-choi@knu.ac.kr (S. Choi).

Abbreviations

ΔV_{air}	change in local air volume
ADI	anisotropic deformation
CNN	convolutional neural network
CT	computed tomography
DIR	deformation image registration
DL	deep learning
DSC	Dice score
DVF	displacement vector field
EX	expiration
IN	inspiration
J	Jacobian
LRN	lung registration network
SSTVD	sum of squared tissue volume difference
STL	spatial transformer layer
STN	spatial transformer network
TRE	target registration error

its expensive computations and nonlinearity of large deformations, and thus, failed to meet the clinical requirements of real-time registration.

The recent significant advancements in learning-based algorithms enable these algorithms to manage many tasks involved in medical image processing, including reconstruction, segmentation, and registration. Based on a pair of input images, learning-based registration methods attempt to learn pixel-wise spatial correspondence by learning the parameters underlying convolution neural networks (CNNs). In contrast to the traditional registration method, the trained model is used to skip an iterative optimization of a similarity metric to produce the displacement vector field (DVF) of the test subject. Therefore, deep learning (DL) is a promising research direction for solving registration problems in clinical practice.

To date, several studies have leveraged learning-based techniques to improve medical image registration. These studies described supervised registration methods [12–16], trained using labeled reference deformations. Eppenhof et al. [16] used a combination of synthetic random transformations (an additional affine preregistration step) and the U-Net model [17] to estimate the deformation part of the transformation in a supervised manner. Sokooti et al. [15] proposed a multistage CNN to minimize the difference between ground-truth DVFs and artificial DVFs. Their algorithm has been proven to be accurate in registering lung images on the DIR-Lab dataset. The supervised methods require ground-truth DVF or realistic synthetic deformations that are extracted by automatic or semiautomatic methods. Thus, the results might be highly sensitive to the DVF quality obtained from ground-truth images and require significant effort to acquire suitable quality results. Consequently, supervised methods have not been widely used for medical imaging registrations.

In contrast, unsupervised registration methods [18–21] do not require ground-truth deformation images and rather focus on minimizing the difference between warp-moving and fixed images and smoothing registered deformation fields by defining the loss functions. De Vos et al. [22] proposed deep unsupervised image learning registration using normalized cross-correlation as a loss function. Max et al. [14] proposed a spatial transformer network (STN), in which their spatial transformation parameters could be learned without labeled images by a specifically designed loss function. Many studies have recently inserted STN into existing CNNs to perform registration and reconstruction tasks [23–25]. These methods are very fast in testing subjects and can learn automatically from data; however, the predicted deformations cannot be guaranteed, thereby causing noisy DVF. Fu et al. [26] presented LungRegNet to register 4D computed tomography (CT) lung images rapidly and accurately. This network consists of CoarseNet

for deforming moving images and FineNet for training. Before training, they developed vessel-enhanced images as a preprocessing step to increase registration accuracy. The majority of state-of-the-art unsupervised registration algorithms [24,27], such as the VoxelMorph [24] devised by Balakrishnan et al., use encoder-decoder architectures based on U-Net to predict DVFs. VoxelMorph was trained and tested on a substantially faster brain magnetic resonance imaging dataset with an accuracy comparable to that of iteration-based methods. Despite its high accuracy and efficiency, VoxelMorph aligned and evaluated any input image against a standard image (atlas image). This is not convenient for image registration where standard images are unavailable, particularly lung deformation images in the present study.

To overcome the aforementioned disadvantages in medical registration, large deformations have been solved using multilevel strategies and iteratively trained networks [16,22,28,29]. Jiang et al. [28] designed a multiscale DIR framework with unsupervised joint training of a CNN for DVF estimation. Zhao et al. proposed a recursive cascaded network [30]. They stacked multiple CNNs to predict the velocity fields in an unsupervised manner and refined the registration result progressively. Another similar method [31] devised by Zhao et al., assigned inputs to each cascade subnetwork and added an invertibility loss to the network. Many learning-based networks are designed as pyramid strategies to perform coarse-to-fine estimation of the deformation field [32, 33] and have achieved competitive accuracy comparable to that of VoxelMorph. Although these learning-based methods have advantages in terms of time and generalization with different types of input images, they still cannot compete with conventional methods in terms of accuracy for lung CT benchmarks (distance <1 mm [34]) with large displacements.

In lung registration, many regions exist where one image has been “folded” over itself by a non-invertible transformation, leading to a negative Jacobian determinant of the deformation field in these regions [8]. These spatial folds can cause registration errors in clinical applications. Christensen and Johnson [35] used the Jacobian value to check for the folding of the deformation and monitored this value during consistent image registration to ensure a one-to-one transformation. Rohlfing et al. [36] penalized the deviations of the Jacobian value from 1 to develop a volume-preserving non-rigid registration algorithm. The frequency of Jacobian-induced errors has limited the adoption of neural network methods for medical image registration.

To address these problems, we propose an unsupervised end-to-end learning-based method—the lung registration network (LRN)—to register CT lung datasets acquired at two different volumes. In the test phase, the LRN outputs consist of a deformation field based on a given lung image pair. Similar to unsupervised learning methods, it does not require labeled data and can satisfy the clinical requirements of real-time registration. The proposed approach is discussed below.

- An LRN network for extracting DVF is proposed based on cycle-consistent training, with an enhanced architecture (MultiResUNet) rather than the popular U-Net-based network [24] (VoxelMorph).
- The LRN network takes a pair of full lung volumes as input and predicts the DVF with a loss function that includes three components: similarity loss based on the sum of squared tissue volume difference (SSTVD), smoothing loss based on L_2 regularization, and explicit anti-folding regularization by a Jacobian determinant.
- The LRN was trained with multiple datasets and assessed using many evaluation metrics (target registration error (TRE), Dice score, and the number of folding voxels). First, 705 pairs of lung CT volumetric datasets were collected from two hospitals in South Korea and trained. Each pair of lung volumes was acquired at full inspiration (IN) and end-expiration (EX). IN provides information related to airway and parenchyma structures, whereas EX provides information related to air trapping [8]. In the testing phase, ten pairs of CT lung images from different public DIR-Lab [37] datasets were used to evaluate the LRN’s response to the registration accuracy, dice score

of spatial overlap, anti-folding performance, and computational time via state-of-the-art learning-based methods (VoxelMorph) and traditional methods (SSTVD) [8].

2. Methods

In image registration, moving image X is deformed to match fixed image Y according to a dense deformation vector field φ to determine a spatial transformation that can match these two images [20]. In this study, the optimal deformation field $\varphi = MRU_\theta$ was estimated using a CNN for accurate image registration.

Fig. 1 presents an overview of the proposed LRN method. The LRN model was trained using a cycle-consistent design [38]. First, a pair of 3-dimensional (3D) images (fixed and moving) was provided in a two-channel 3D image and fed into MultiResUNet. Note that this channel is typically used to collect RGB images. The MultiResUNet model learns parameters θ_1 for the function $MRU_{\theta_1}(X, Y)$ to compute φ_{XY} , which registers a 3D moving volume X to a fixed volume Y . During training, X is warped with the displacement field φ_{XY} using a spatial transformer function to create a warped moving image \tilde{Y} and to evaluate the similarity of \tilde{Y} and Y . Second, the same MultiResUNet model learns parameters θ_2 for a function $MRU_{\theta_2}(\tilde{Y}, X)$ that registers the warped volume \tilde{Y} (now moving) to the moving volume X (now fixed). During training, \tilde{Y} was warped with the corresponding inverted displacement field $\varphi_{\tilde{Y}X}$ using the same spatial transformer function to create \tilde{X} and evaluate the similarity of \tilde{X} and X . Given new fixed and moving images during the test phase, the registration fields were obtained by evaluating $MRU_{\theta_1}(X, Y)$ and $MRU_{\theta_2}(\tilde{Y}, X)$. The expected loss function was minimized using a training dataset to determine the optimal parameters θ_1 and θ_2 . The losses compare \tilde{Y} , Y and \tilde{X} , X and enforce smoothness of φ_{XY} and $\varphi_{\tilde{Y}X}$. In the following sections, the MultiResUNet structure and loss functions are described.

2.1. Spatial transformer layer (STL)

Inspired by the recent success of STN [14], in this study, the LRN model was developed by combining STL with MultiResUNet networks for computing warped moving images [39]. Here, the STL was inserted into the MultiResUNet registration network to estimate warped moving images \tilde{Y} and \tilde{X} that are the composition of the deformation field and the moving image. The position ($p' = p + \varphi(p)$) of each voxel p , with $p = (x,$

$y, z)$ in the moving image was computed in the space of the fixed image. The sampling grid and sampler were the two main components of STL (Fig. 1). The predicted deformation field φ using the MultiResUNet model was used to transform the regular spatial grid into a sampling grid using a grid generator. A sampling grid was used to spatially warp the moving image inside MultiResUNet by linear interpolation. As new sampling locations can be non-integral, the values at the eight neighboring voxels were linearly interpolated in the x , y , and z directions. The gradients were computed, and the errors were backpropagated during the optimization of the MultiResUNet model. In addition, its fully differentiable property enabled the end-to-end training of the LRN registration network [20].

2.2. Unsupervised image registration network–MultiResUNet network

Fig. 2 shows the MultiResUNet-CNN network used in the present study for estimating lung deformation fields. This network includes many MultiResolution (MultiRes) blocks and Residual (Res) paths that differ from those of the original U-Net, based on the replacement of the sequences of the two convolution layers in the U-Net architectures with MultiRes blocks and Res paths, as shown in Fig. 2. In the MultiResUNet architecture, the MultiRes block is placed before each max-pooling and transposed convolution layer. The MultiRes block is a sequence of $3 \times 3 \times 3$, $5 \times 5 \times 5$, and $7 \times 7 \times 7$ convolution filters (Inception blocks) in parallel to generate feature maps that can analyze images at different scales (Fig. 2). Instead of combining the encoder feature maps with the decoder feature using skip connection and concatenation, the encoder features were passed through a sequence of convolutional layers. Res paths were included to combine high- and low-level features [39].

2.3. Training loss functions

The MultiResUNet parameters are updated by minimizing the combination loss. This unsupervised loss function L_{us} comprises three components:

$$\begin{aligned} L_{us}(X, Y, \varphi) = & L_{sim}(Y_V, \tilde{Y}_V) + L_{sim}(X_V, \tilde{X}_V) + \lambda_1[L_{smooth}(\varphi_{XY}) + L_{smooth}(\varphi_{\tilde{Y}X})] \\ & + \lambda_2[L_{J_det}(\varphi_{XY}) + L_{J_det}(\varphi_{\tilde{Y}X})], \end{aligned} \quad (1)$$

where λ_1, λ_2 are the regularization hyperparameters.

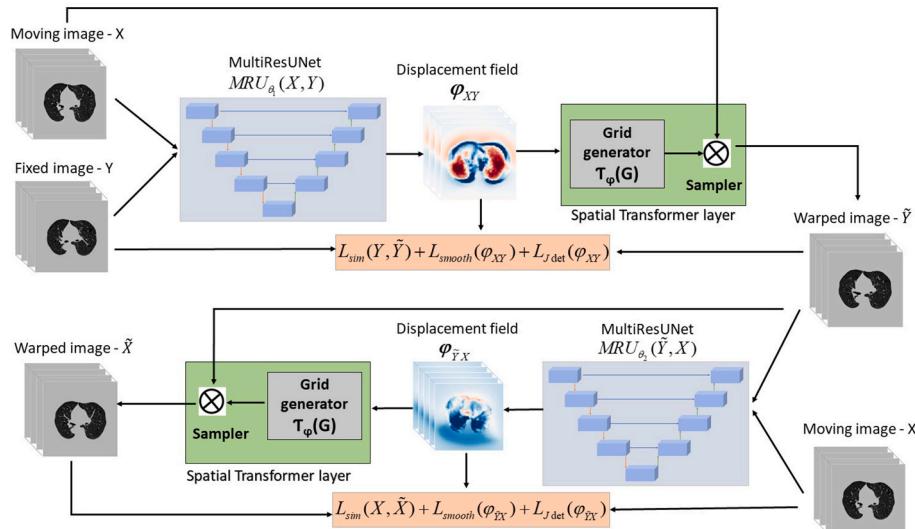


Fig. 1. Overview of the proposed method.

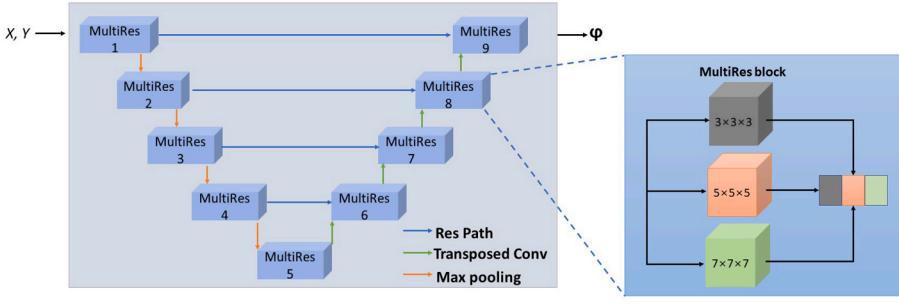


Fig. 2. Architecture of the MultiResUNet structure, including MultiRes blocks and Res paths. The two convolution layers in the U-Net structure were replaced with MultiRes blocks to generate feature maps (left). Instead of using a plain shortcut skip connection, Res paths were proposed to combine the encoder feature maps with the decoder feature. The Inception block was used by employing $3 \times 3 \times 3$, $5 \times 5 \times 5$, and $7 \times 7 \times 7$ convolutional filters in parallel and concatenating the generated feature maps to reconcile spatial features from different context sizes (right).

2.3.1. Similarity loss

The first term L_{sim} is based on SSTVD and is used to correct the intensity change of lung CT images during breathing by measuring the image similarity between warped moving and fixed images. The SSTVD approach extends the sum square difference and assumes that tissue volume is preserved throughout the respiratory cycle by including air and tissue volume information [8,21]. The Hounsfield (HU) of the lung CT images includes two components: tissue ($HU_{tissue} = 55$) and air ($HU_{air} = -1000$). The tissue volume of the fixed image (X_V) and moving image (Y_V) associated with fixed intensity (X_I) and moving intensity (Y_I), respectively, can be estimated as follows:

$$X_V = \frac{X_I - HU_{air}}{HU_{tissue} - HU_{air}} = \frac{X_I + 1000}{1055}, Y_V = \frac{Y_I - HU_{air}}{HU_{tissue} - HU_{air}} = \frac{Y_I + 1000}{1055}. \quad (2)$$

The Jacobian determinant matrix is the derivative of the displacements, which estimates the local volume change between the two images as follows:

$$J(\varphi(\mathbf{p})) = \begin{pmatrix} \frac{\partial \varphi_x(p)}{\partial x} & \frac{\partial \varphi_x(p)}{\partial y} & \frac{\partial \varphi_x(p)}{\partial z} \\ \frac{\partial \varphi_y(p)}{\partial x} & \frac{\partial \varphi_y(p)}{\partial y} & \frac{\partial \varphi_y(p)}{\partial z} \\ \frac{\partial \varphi_z(p)}{\partial x} & \frac{\partial \varphi_z(p)}{\partial y} & \frac{\partial \varphi_z(p)}{\partial z} \end{pmatrix}. \quad (3)$$

The tissue volume between the fixed and moving images is related using $X_V = |J(\varphi)| \cdot Y_V$. The SSTVD loss function is defined as follows:

$$L_{sim}(Y_V, \tilde{Y}_V) = L_{SSTVD}(Y_V, \tilde{Y}_V) = \sum_{p \in \Omega} (Y_V(\mathbf{p}) - |J(\varphi(\mathbf{p}))| \cdot \tilde{Y}_V(\mathbf{p}))^2, \quad (4)$$

where \tilde{Y}_V is computed by warping the moving tissue volume X_V with φ . Similar approximations, $L_{sim}(X_V, \tilde{X}_V) = L_{SSTVD}(X_V, \tilde{X}_V) = \sum_{p \in \Omega} (X_V(\mathbf{p}) - |J(\varphi(\mathbf{p}))| \cdot \tilde{X}_V(\mathbf{p}))^2$.

2.3.2. Regularization loss

The second term L_{smooth} encourages the smoothness of the predicted displacement field φ using a regularization function (L_2) on the spatial gradients of φ using the differences between neighboring voxels.

$$L_{smooth} = L_{smooth}(\varphi_{XY}) + L_{smooth}(\varphi_{YX}) = \sum_{p \in \Omega} \|\nabla \varphi_{XY}(\mathbf{p})\|^2 + \sum_{p \in \Omega} \|\nabla \varphi_{YX}(\mathbf{p})\|^2, \quad (3)$$

where $\nabla \varphi(\mathbf{p}) = \left(\frac{\partial \varphi(\mathbf{p})}{\partial x}, \frac{\partial \varphi(\mathbf{p})}{\partial y}, \frac{\partial \varphi(\mathbf{p})}{\partial z} \right)$.

2.3.3. Jacobian loss

The model's registration accuracy may be significantly reduced by global regularization (L_2), particularly if a significant weight is provided for regularization [40]. To guarantee a topology preservation transformation, these regularizations are insufficient. The second

regularization term L_{Jdet} penalizes the local region with a negative Jacobian determinant on the estimated DVFs, in which folding exists on the warped moving images [40].

$$L_{Jdet} = \frac{1}{N} \sum \sigma(-|J(\varphi(\mathbf{p}))|) = \frac{1}{N} \sum \sigma(-|\det(\mathbf{D})|), \quad (4)$$

where $\det(\mathbf{D}) = J(\varphi(\mathbf{p}))$, N is the total number of elements in $\det(\mathbf{D})$, $\sigma(\cdot) = \max(0, \cdot)$ is the rectified linear unit (ReLU) activation function, and $\det(\mathbf{D})$ is the determinant of the Jacobian matrix at \mathbf{p} . A positive Jacobian (J) value at point $\mathbf{p} \in |J(\varphi)|$ indicates that the displacement field preserves the orientation in the vicinity of \mathbf{p} . If $\mathbf{p} \in |J(\varphi)|$ is negative, the displacement field at \mathbf{p} reverses the orientation in the neighborhood of \mathbf{p} ; thus, the one-to-one mapping is lost and should be penalized [40]. The Jacobian value of the displacement reflects the local contraction or expansion [8].

3. Datasets and experiment

3.1. Datasets for training and testing

3.1.1. Training datasets

All data were collected and approved according to the guidelines and regulations of the Institutional Review Board of the participating centers. A total of 705 subjects (397 subjects exposed to cement dust, 308 normal subjects) were enrolled and underwent CT scans at the Kangwon National University Hospital (KNUH) and Jeonbuk National University Hospital (JNUH). The details of the demographic information of the training dataset are shown in Table 1. The gender, age, height, and weight were significantly different between two groups ($P < 0.001$). The acquisition and reconstruction parameters of the two CT scanners were obtained by referring to the study conducted by Cho et al. [41]. Both CT scanners featured similar specifications and imaging protocols and used the same filtered back-projection reconstruction kernel (Siemens Definition Flash 128 slices B30f and Siemens Definition AS 64 slices B35f) [42]. Two volumetric scans of each subject were acquired during breath-holds near IN and EX in the same scan session. Each volumetric dataset contained 500–800 image sections with a sectional spacing of 0.5–0.7 mm and a reconstruction matrix of 512×512 px. The in-plane pixel spatial resolution was approximately 0.6×0.6 mm 2 . For each

Table 1

Demographics for cement dust-exposed and normal subjects for the training dataset.

Demographics	Normal subjects ($n = 308$)	Cement dust-exposed subjects ($n = 397$)	P-value
Male/Female, n	132/176 (42.86%/57.14%)	301/96 (75.82%/24.18%)	<0.001
Age, years	51.38 (15.18)	73.10 (7.37)	<0.001
Height, cm	162.06 (9.91)	159.65 (9.09)	<0.001
Weight, kg	64.34 (12.29)	59.13 (10.41)	<0.001

Values are presented as mean (standard deviation (SD)). P-value is calculated based on Wilcoxon paired tests for statistical comparison.

subject, automatic lung segmentation was performed using Apollo software (VIDA Diagnostics, Coralville, IA).

3.1.2. Testing datasets

The method was tested using ten pairs of CT scans from the public DIR-Lab [37,43] dataset to evaluate the accuracy of the proposed registration method. DIR-Lab included the 4DCT datasets, which contained data from the 4DCT scan sequences of ten patients, with each consisting of five 3DCT images covering the respiratory cycle from end IN to end EX breath-hold CT scans. For each pair of IN-EX scans, a set of 300 corresponding landmarks was provided by an expert using a semi-automatic tool [37,43,44]. All subjects in the testing dataset were segmented using lobes and airway segmentation modules of the 3D slicer chest imaging platform (CIP) (Brigham and Women's Hospitals) [45]. The background of the lung images is complicated because there are various anatomical structures in the abdomen. Background noise from other visceral organs or vessels influences the lung registration process negatively. For such tasks, lung segmentation [26,46] is a prerequisite for accurate image registration, as it can focus registration on the lung region and avoid undesired alignment of artifacts. Therefore, the lung was first segmented using Apollo and CIP for registration in the training and testing phases, respectively. Subsequently, the lung was registered based on the segmentation of the 3DCT images.

3.2. Experiment

The registration accuracy of a pair of images was evaluated by TRE as the average Euclidean distance between a landmark in the warped moving image and its corresponding landmark in the fixed image [29] as shown in Equation (5):

$$TRE = \|l_m + \varphi(l_m) - l_f\|, \quad (5)$$

where l_f and l_m are the corresponding landmarks on the fixed and moving images, respectively, and $\varphi(l_f)$ denotes the estimated DVF. In addition, the dice similarity coefficient (DSC) was used to measure the spatial overlap of the anatomical map between the fixed volume D_f and warped moving volume $D_{m\circ\varphi}$. The dice score ranged from 0 to 1, with a value of 1 indicating a greater similarity between D_f and $D_{m\circ\varphi}$ (Equation (6)):

$$DSC = \frac{2|D_f \cap D_{m\circ\varphi}|}{|D_f| + |D_{m\circ\varphi}|}. \quad (6)$$

The regional air volume change (ΔV_{air}), volume change (measured by J), and anisotropic deformation index (ADI) quantify lung deformation [47]. With the eigenvalues ($\lambda_1, \lambda_2, \lambda_3$), J and the ADI were calculated as follows:

$$\Delta V_{air}(\mathbf{p}) = V_{air}^f(\mathbf{p}) - V_{air}^{m\circ\varphi}(\mathbf{p}), \quad (7)$$

$$J = \lambda_1 \lambda_2 \lambda_3, \quad (8)$$

$$ADI = \sqrt{\left(\frac{\lambda_1 - \lambda_2}{\lambda_2}\right)^2 + \left(\frac{\lambda_2 - \lambda_3}{\lambda_3}\right)^2}. \quad (9)$$

ΔV_{air} was subsequently normalized by the respective medians of the same subject [47], denoted by ΔV_{air}^* . In this study, the left upper, left lower, right upper, right middle, and right lower lobes are denoted as LUL, LLL, RUL, RML, and RLL, respectively. Additionally, the number of folding voxels could be counted using the Jacobian determinant of the displacement field for each method.

LRN and VoxelMorph were trained on all pairs of images from the training dataset, and their predicted deformations were examined using pairs of images from the testing dataset. Although our LRN model could analyze image pairs of arbitrary size, all 3DCT images were resized to $160 \times 192 \times 224$ voxels to align them with the GPU memory and match

the input size of VoxelMorph. The CT image intensity was normalized to the range [0, 1]. Both networks were trained on the training set with the same hyperparameters, using Adam optimizer with a learning rate of 10^{-4} , 1500 epochs. The networks were implemented in Python 3.7 and TensorFlow on a computer equipped with NVIDIA Tesla V100 PCIe 32 GB of memory. The hyperparameter $\lambda_1 = 2, \lambda_2 = 10^{-3}$ must be manually chosen. Analysis of variance (ANOVA) along with Tukey's post-hoc tests was performed for analyzing significance, and $P < 0.05$ was considered significant.

4. Result

4.1. Comparison cycle-consistent training and direct unsupervised training

Table 2 and Fig. 3 quantitatively present the performance of the proposed unsupervised method (LRN) with and without cycle-consistent training. For LRN with and without cycle-consistent training, approximately 41 h and 107 h are required by NVIDIA Tesla V100 PCIe, respectively. For visualization purposes, the axial lung CT slices of two subjects from moving and fixed images and the resulting warped and Jacobian images of each method are presented in Fig. 4. TRE was used as an evaluation metric for all subjects and structures in the DIR-Lab test set (Table 2). Fig. 3 exhibits the distribution of landmark distances along the z-axis (from apex to base) before and after registration for each method. The initial TREs increase as landmarks approach the base (Fig. 3) because the lung deformation is mainly driven by the diaphragm. The landmark distances decrease with both the LRN with and without cycle-consistent training methods. Fig. 3 also shows the large deviation in errors in the landmarks for the LRN without the cycle-consistent method. The LRN with cycle-consistent training was slightly better than that without cycle-consistent training, with TREs of 1.78 ± 1.56 mm and 1.86 ± 1.98 mm on the DIR-Lab dataset, respectively. Fig. 4 displays an appropriate alignment of the overall shapes of the lungs and structures within the lungs for both methods. In addition, the warped and Jacobian images of the moving and fixed images after registration were similar for the two methods (Fig. 4).

4.2. Target registration errors (TREs) on the DIR-lab dataset

Table 3 and Fig. 5 quantitatively show the performance of the proposed unsupervised method (LRN) along with the VoxelMorph and SSTVD traditional registration algorithms. For visualization purposes, the axial and coronal lung CT slices of a subject from the moving, fixed, and resulting warped images of each method are presented in Fig. 6. DSC and TREs were used as evaluation metrics for all subjects and structures in the DIR-Lab test set. Fig. 5 displays the distribution of the landmark distances along the z-axis after the registration of each method. All three methods significantly reduced TREs on the DIR-Lab dataset after image registration (Fig. 5). Fig. 5 also indicates that the TREs have a large deviation for the LRN and VoxelMorph methods. The performance of LRN was compared to that of SSTVD and was found to be better than the U-Net-based model, that is, VoxelMorph, with TREs of 1.78 ± 1.56 , 1.66 ± 1.49 , and 2.43 ± 2.43 on the DIR-Lab dataset, respectively. In

Table 2

Target registration errors (lower is better) between the fixed and warped moving images of the DIR-Lab dataset for the proposed lung registration network with and without cycle-consistent training.

DIR-Lab dataset	TRE before Registration (mm)	TRE after registration of LRN (mm)		P-value (ANOVA)
		With cycle-consistent training	Without cycle-consistent training	
Average	8.46 ± 6.58	1.78 ± 1.56	1.86 ± 1.98	0.58

LRN, lung registration network; TRE, target registration error.

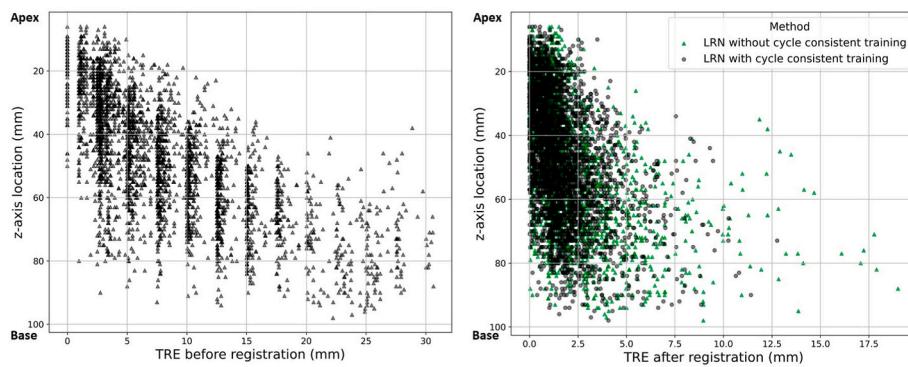


Fig. 3. Scatter plots showing the distribution of target registration errors (TREs) along the z-axis before and after registration of lung registration network (LRN) with and without cycle-consistent training.

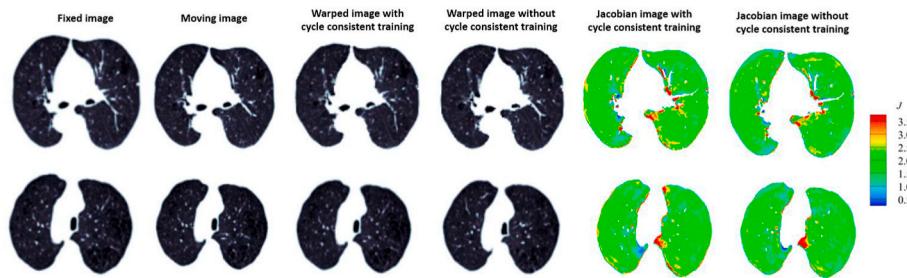


Fig. 4. Comparison of registration results obtained with and without cycle-consistent training. No folding in the Jacobian images exists.

Table 3

Target registration errors (lower is better) between the fixed and warped moving images of the DIR-Lab dataset for the proposed lung registration network, the sum of squared tissue volume difference, and VoxelMorph methods.

Dataset	TRE before Registration (mm)	TRE after registration (mm)			P-value (ANOVA)	*P-value between LRN and SSTVD
		LRN	SSTVD	VoxelMorph		
DirLab01	3.89 ± 2.78	1.14 ± 0.71	1.12 ± 1.04	1.16 ± 0.80	0.85	0.96
DirLab02	4.34 ± 3.90	1.04 ± 1.13	1.07 ± 0.73	1.09 ± 0.77	0.79	0.91
DirLab03	6.94 ± 4.05	1.44 ± 0.84	1.25 ± 0.58	1.60 ± 1.02	<0.001*†‡	0.01
DirLab04	9.83 ± 4.86	1.54 ± 1.20	1.41 ± 0.91	2.46 ± 1.85	<0.001†‡	0.48
DirLab05	7.48 ± 5.51	1.60 ± 1.63	1.67 ± 1.49	1.96 ± 1.66	0.01†	0.85
DirLab06	10.89 ± 6.97	2.34 ± 1.29	2.03 ± 1.58	2.85 ± 2.13	<0.001†‡	0.07
DirLab07	11.03 ± 7.43	1.80 ± 0.90	1.99 ± 1.37	3.48 ± 2.75	<0.001†‡	0.42
DirLab08	14.99 ± 9.01	3.76 ± 2.52	3.12 ± 2.15	5.17 ± 4.20	<0.001*†‡	0.03
DirLab09	7.92 ± 3.98	1.62 ± 1.19	1.56 ± 1.05	2.26 ± 1.56	<0.001†‡	0.84
DirLab10	7.30 ± 6.35	1.57 ± 1.54	1.37 ± 1.94	2.29 ± 2.36	<0.001†‡	0.43
Average	8.46 ± 6.58	1.78 ± 1.56	1.66 ± 1.49	2.43 ± 2.43	<0.001*†‡	0.04

Values are mean ± SD. ANOVA tests with Tukey's post-hoc tests of TREs among LRN, SSTVD, and VoxelMorph. *P < 0.05 for LRN and SSTVD, †P < 0.05 for LRN and VoxelMorph, and ‡P < 0.05 for SSTVD and VoxelMorph. LRN, lung registration network; SSTVD, the sum of squared tissue volume difference; TRE, target registration error.

addition, Fig. 6 demonstrates the overlapping images of moving and fixed images before registration and between fixed and warped moving images after registration attained from the traditional SSTVD, VoxelMorph, and LRN methods. The intensity differences between the fixed and warped moving images were observed to be greater for VoxelMorph than those for SSTVD and LRN, indicating an improved overall lung alignment for SSTVD and LRN. The red arrows in Fig. 6 indicate a suitable alignment structure in the lung for the SSTVD and LRN methods. Although both approaches exhibited a suitable alignment of the overall shapes of lungs and structures within the lungs, blue arrows in Fig. 6 show an extent of misalignment between the fixed and warped moving images, which was observed in VoxelMorph but not in SSTVD and LRN.

4.3. Spatial characteristics of normalized air volume change (ΔV_{air}^*), volume change (J), and anisotropic deformation index (ADI)

Figs. 7 and 8 show the lobar normalized air volume changes excluding the tissue volume (ΔV_{air}^*), lung volume change including tissue volume (J), and ADI of each registration method. A mixed between- and within-group ANOVA was performed to evaluate three dependent variables (ΔV_{air}^* , J, and ADI) for three independent groups (LRN, SSTVD, and VoxelMorph) and five lung regions (LUL, LLL, RUL, RML, and RLL). The differences in ΔV_{air}^* , J, and ADI among the three methods were not significant. ΔV_{air}^* ($P = 0.97$) and ADI ($P = 0.98$) were similar for all three methods. For the within-group variable "lung regions," three dependent variables indicated that none of the lobes had different characteristics of ΔV_{air}^* , J, and ADI (Table 3). The folding results are shown in Jacobian images after the image registration of each

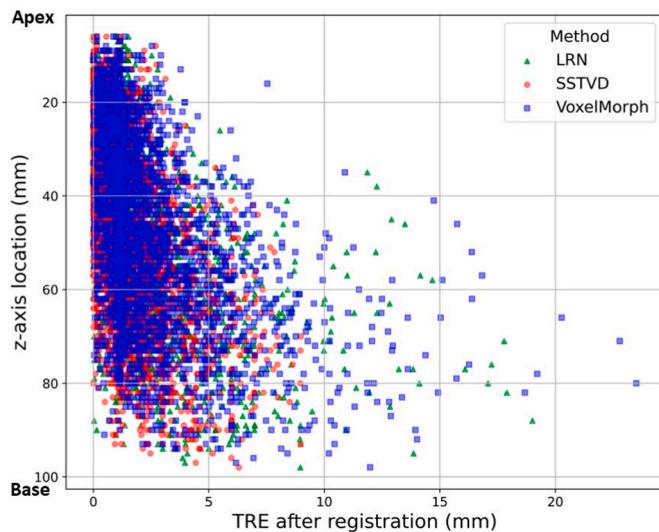


Fig. 5. Scatter plot showing the distribution of TREs along the z-axis after registration of each method.

method. Unlike the VoxelMorph method, the proposed and SSTVD methods do not fold in the predicted deformations, as shown in Table 4 and Fig. 8.

4.4. Effect of jacobian loss

Gradients exist with larger J and ADI in the lung border of learning-based methods compared to those of SSTVD (Fig. 8). Table 5 presents the results of the dice scores, used to quantify the volume overlap between the two volumes. These dice scores were all $<0.68 \pm 0.19$. This result can be explained based on pairwise registration (instead of registration to an atlas [24]) and all region of interests (ROIs), which is finer than VoxelMorph with a defined number of ROIs. The proposed models achieve comparable dice measures (0.65 ± 0.24) to SSTVD (0.68 ± 0.19) for all structures ($P = 0.96$), performing significantly better than VoxelMorph (0.51 ± 0.27). In particular, the proposed method with the additional Jacobian loss function achieved no folding voxels (compared to VoxelMorph), as shown in Table 5.

4.5. Running time analysis

The running time for each algorithm to register a pair of lung CT images using an NVIDIA GTX 2080Ti GPU is reported in Table 5. SSTVD requires approximately eight or more minutes for one pair of lung CT images. VoxelMorph and the proposed methods are considerably faster on average upon using a GPU than traditional methods.

5. Discussion

An unsupervised learning architecture, that is, LRN, without any ground-truth deformation, was developed for the training of 3D medical image registration. This LRN network was trained with a combined loss function with an anti-folding part, which penalizes the spatial locations of non-invertible deformation using a negative Jacobian determinant. LRN was experimentally compared with VoxelMorph and SSTVD using pairwise registration on the DIR-Lab dataset. LRN achieved state-of-the-art results in 3D lung registration in terms of accuracy and registration speed without any folding. The proposed method is particularly advantageous when the volume difference between the two volume CT lung images is large.

5.1. Sum of squared tissue volume difference (SSTVD) traditional registration method and the folding of the deformation field

A traditional image registration algorithm requires several components that are combined with the cost function (also known as the loss function in learning-based methods), multiresolution/multiscale strategy and coordinate transformation model [8,48] as shown in Fig. 9. In this study, we used SSTVD as a baseline method to compare the warped moving images created by the DL methods. SSTVD is based on quantitative CT measures and aims to minimize the local tissue volume difference within the lungs between the matched regions [8]. SSTVD achieved satisfactory performance on various datasets. However, for each pair of unseen lung CT images to be registered, SSTVD and most conventional registration methods iteratively optimize the cost function from scratch and manually tune parameters, which is time-consuming and neglects the underlying registration patterns shared by all images in the same dataset [20]. In this study, a fully convolutional image-to-image registration architecture can learn an optimal DVF from the training dataset and quickly output the DVF for each pair of images. This overcomes the above-mentioned drawbacks and does not involve human participation in traditional methods, particularly in real-time

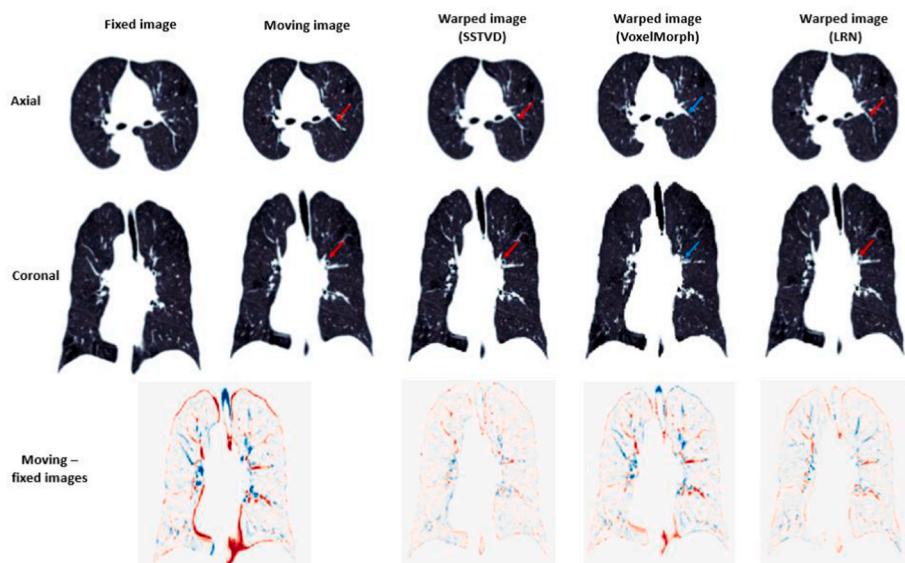


Fig. 6. Axial and coronal lung computed tomography slices of registration results of a subject from fixed, moving, and resulting warped images of each method, respectively (first and second rows). The suitable alignment between the vessel trees for the LRN and SSTVD methods (indicated by the red arrow) and a misalignment of the airway for the VoxelMorph method (indicated by the blue arrow) is noticeable. The overlap images-intensity difference images of moving and fixed images before registration and between fixed and warped moving images after registration from traditional (SSTVD), VoxelMorph, and the proposed LRN method (third row).

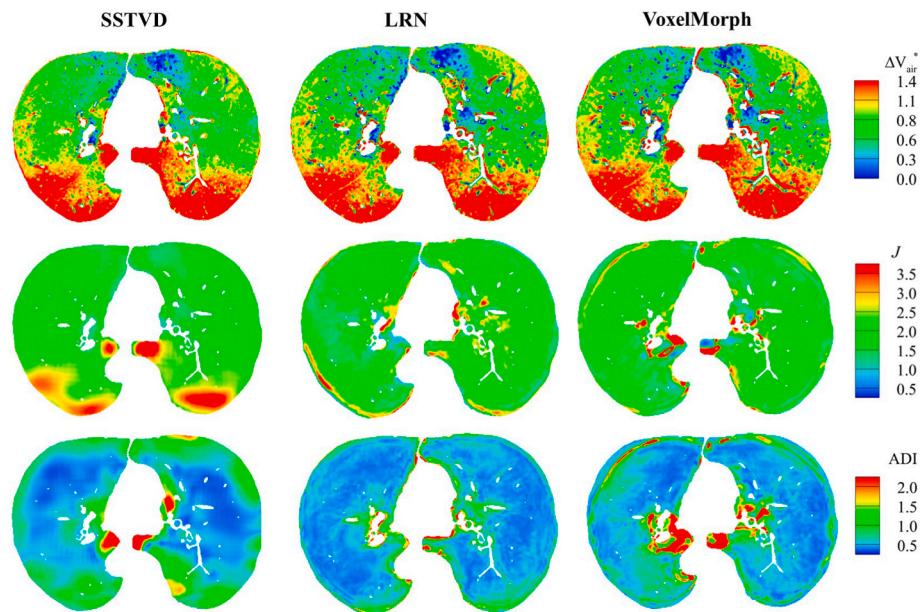


Fig. 7. Regional distribution of air volume change ΔV_{air}^* , J , and ADI obtained via each registration method from the ventral (top) to the dorsal region (bottom).

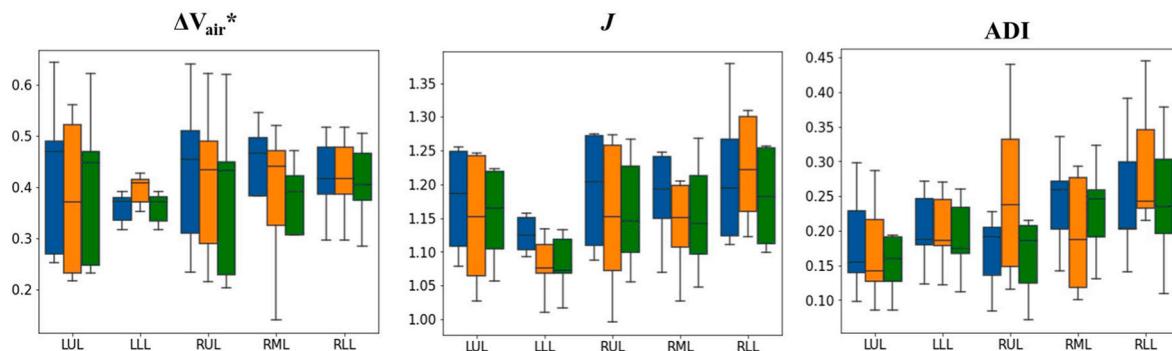


Fig. 8. Mean and standard deviation of normalized air volume change (ΔV_{air}^*), volume change (J), and anisotropic deformation index (ADI) of the VoxelMorph, lung registration network (LRN), and the sum of squared tissue volume difference (SSTVD) registration methods. The values are presented as boxes (bottom line: 25th percentile; middle line: median; top line: 75th percentile) and whisker plots (bottom line: 5th percentile; top line: 95th percentile). LUL, left upper lobe; LLL, left lower lobe; RUL, right upper lobe; RML, right middle lobe; RLL, right lower lobe.

Table 4

Mixed (between- and within-group) ANOVA test was performed on LRN, SSTVD versus VoxelMorph (between), and five lobes (repeated measures) as a grouping and within a variable, respectively.

ANOVA (F-Test, P-value)	LRN, SSTVD, vs. VoxelMorph		LRN vs. SSTVD		VoxelMorph vs. SSTVD	
	Groups	Lung regions	Groups	Lung regions	Groups	Lung regions
ΔV_{air}^*	0.97	1.0	0.94	0.93	0.83	1.0
J	0.86	1.0	0.77	0.98	0.58	0.98
ADI	0.98	0.90	0.89	1.0	0.90	0.99

ΔV_{air}^* , air volume change; J , volume change; ADI, anisotropic deformation; LRN, learning registration network; SSTVD, the sum of squared tissue volume difference.

applications. Furthermore, the quantitative results obtained for LRN using dice, TRE, J , and ADI were close to those obtained using SSTVD.

5.2. Effect of the jacobian loss function

The loss function is one of the most important components of an

Table 5

Average and standard deviation of the Dice score (higher is better), running time (lower is better), and the number of folding voxels for each registration method.

Variables	LRN	SSTVD	VoxelMorph	P-value (ANOVA)	*P-value (post-hoc tests) between LRN and SSTVD
Dice score	0.65 \pm 0.24	0.68 \pm 0.19	0.51 \pm 0.27	0.25	0.96
Time (s)	1.67 \pm 0.02	463 \pm 15.33	1.58 \pm 0.01	<0.001	0.87
Folding (voxel)	0.00 \pm 0.00	0.00 \pm 0.00	3233 \pm 2156	<0.001†‡	–

Values are mean \pm SD. ANOVA tests with Tukey's post-hoc tests of the Dice score, computational time, and number of folding voxels among LRN, SSTVD, and VoxelMorph. * $P < 0.05$ for LRN and SSTVD, † $P < 0.05$ for LRN and VoxelMorph, and ‡ $P < 0.05$ for SSTVD and VoxelMorph. SSTVD is the sum of squared tissue volume difference.

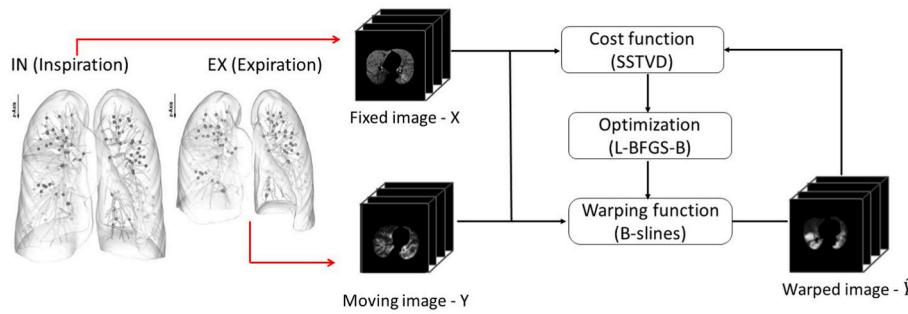


Fig. 9. Flow chart depicting the SSTVD traditional registration method. The two input images pass through four registration steps until an optimizer stop condition is reached. SSTVD is the sum of squared tissue volume difference. L-BFGS-B: optimize displacement of control grids. The tissue volume is assumed to remain the same between two inflation levels [8].

unsupervised registration model. Herein, a regularization term (L_2) was adopted to penalize the undesired deformations. As changes in voxel intensity vary from the apex to base and ventral to dorsal, this regularization loss may not be sufficient to capture the non-uniform expansions of the lung [8]. The changes in tissue volume owing to blood volume and lung motion were partly against the smoothness of the deformation fields. The folding in the DVF can be calculated based on the Jacobian values. A positive value of the Jacobian is satisfied if the transformation is invertible. The results of this study suggest the importance of considering such intensity changes in lungs associated with inflation while registering CT lung images using the Jacobian loss criterion.

This proposed loss function is also used to minimize the local tissue volume difference within the lungs during respiration. The difference in local tissue volume was attributed to two factors. First, the change in regional volume was calculated from the Jacobian value of the deformation. Second, the change in the fractional tissue content within a region due to respiration was estimated using SSTVD [8]. To evaluate the effectiveness of the additional Jacobian loss functions, registrations were performed for testing lung CT pairs acquired near IN and EX from the DIR-Lab dataset. The results demonstrate that LRN produces smaller average TREs than those of VoxelMorph and slightly larger errors than those of SSTVD. VoxelMorph used only L_2 regularization loss to smooth the displacement fields, which may cause more regions to fold, leading to lower quantitative accuracy. Visual inspection showed that the LRN also improved the alignment of structures within the lungs.

Moreover, cycle-consistent training contributes to the reduction of negative Jacobian determinants [38]. Without cycle training, the average TRE of the LRN was 1.87 ± 2.41 mm (Table 2). The advantage of this cycle is that it adds the task of recovering the moving image from its already predicted image during training without modifying the backbone architecture, such as the inverse consistent network [49]. The cycle-consistent training only uses (X, Y) in contrast to bi-directional registrations [49] that consider both pairs (X, Y) and (Y, X) as inputs.

5.3. Comparison with previous lung learning-based registration studies

The state-of-the-art learning-based methods are compared below. Table 6 shows that the proposed method is comparable to previous DL methods using TRE on the DIR-Lab dataset.

5.3.1. Supervised learning methods

Supervised registration methods [12–16] train using labeled reference deformations. Sokooti et al. [15] proposed a multistage CNN to minimize the difference between ground-truth DVFs and artificial DVFs. Eppenhof et al. [16] used a combination of synthetic random transformations, an additional affine preregistration step, and a U-Net model to estimate the deformation part of the transformation in a supervised manner. Their algorithm is shown to accurately register lung images on the DIR-Lab dataset with TREs of 1.86 ± 2.12 and 2.17 ± 1.89 mm,

Table 6

Target registration errors of learning-based methods on DIR-Lab datasets. Percentage folding was measured by dividing the voxels of the deformation image with a negative Jacobian determinant by the total voxels.

Previous studies	Method	Training dataset	TRE (mm)	Time (s)	Folding (%)/folding (voxel)
Sokooti et al. [15]	Supervised	19 pairs	1.86 ± 2.12	2.20	$0.01\% \pm 0.01\%$
Eppenhof et al. [16]	Supervised	7 pairs	2.17 ± 1.89	0.58 ± 0.07	1.28%
Wang et al. [21]	Unsupervised	2042 pairs	1.35 ± 0.79	0.93 ± 0.03	0.00
De Vos et al. [22]	Supervised + multiresolution	10 pairs	2.64 ± 4.32	0.63	–
Fechter et al. [23]	Unsupervised + multiresolution	31 pairs	1.83 ± 2.35	60–240	0.25%
Zheng et al. [25]	Unsupervised	30 pairs	2.09 ± 1.55	3.50 ± 0.70	59 ± 5 voxels
Fu et al. [26]	Unsupervised + multiscale	10 pairs	1.59 ± 1.58	–	–
Jiang et al. [28]	Unsupervised + multiscale	22 pairs	1.66 ± 1.44	1.40	<0.1%
Yang et al. [29]	Unsupervised + multiscale	1125 pairs	1.53 ± 1.02	1.0 ± 0.07	–
Proposed LRN	Unsupervised	705 pairs	1.78 ± 1.56	1.88 ± 0.02	0.00

respectively. Compared with these patch-based training systems, image-to-image prediction can be performed using end-to-end full CNNs [50], where pixel-wise features are predicted. Although all the aforementioned studies achieved suitable performance, they were trained on synthetic datasets or datasets using the results of conventional methods as the ground truth [20]. However, ground-truth lung deformation images are difficult to collect.

5.3.2. Unsupervised learning methods

To eliminate the need to collect real data with abundant and reliable ground-truth annotations, more unsupervised learning methods are used [21,22,25,26,28]. Fu et al. [26] adopted unsupervised DL for image registration tasks. Despite obtaining suitable results, they used a patch-based learning system and relied on other feature-based registration approaches. Wang et al. [21] recently achieved better TRE on the DIR-Lab dataset by including the sum of squared vessellness measure

difference loss function between the deformed moving and fixed images. The findings of these studies were utilized in the current study to train a model for image-to-image registration in an unsupervised manner. Most current DL studies use U-Net-like architectures to predict DVFs. The LRN with the MultiResUNet model utilizes a multiresolution strategy to extract features at multiscale levels and improve computational efficiency. In addition, many studies, including VoxelMorph, mapped input images into a displacement space (atlas image), and final warp moving images were reconstructed from this atlas image. In lung registration studies, lung atlas images are not available owing to lung motion.

5.3.3. Multiresolution/multiscale learning methods

Recently, certain learning-based studies used multiresolution/multiscale strategies to improve computational efficiency and optimization stability and avoid local minima [22,23,26,28,29,46]. Additionally, a sequence of reduced-resolution versions of the input images was created, which formed a pyramid representation. Registration was subsequently performed at each level of the pyramid from coarse-to-fine resolutions. Moreover, experiments on IN-to-EX lung registration demonstrated the ability of the LRN to predict large deformations. The limitation of previous studies is the number of subjects in the training dataset; most studies used only DIR-Lab to train and test with less than 50 pairs of images. This decreases the robustness of learning-based networks. Herein, the network was trained and tested using datasets from many sources, which provides strong evidence that the LRN can generalize from different scanners and protocols.

A limitation of this study is the rescale preprocessing step of the training dataset owing to the power of the available computing devices. This may have affected the accuracy of the proposed method. Furthermore, the λ_1 and λ_2 regularization hyperparameters were chosen manually based on the experiments. The memory problem, input size, and method for automatically adjusting the hyperparameters will be improved in future studies.

6. Conclusion

In this study, an unsupervised registration method was introduced to align two 3D CT lung images acquired during breath-holds at different lung volumes within 2 s. The results demonstrate that LRN achieves state-of-the-art results on 3D lung registration in terms of both accuracy and registration speed. Consequently, LRN architecture is a promising alternative to U-Net for many clinical registration applications. The warped moving image is effective for linking structures to function and identifying regional changes in the lungs. Moreover, it can be extended to register organs other than the lungs, such as the brain, liver, kidneys, and heart.

Funding

This work was supported by the Korea Ministry of Environment (MOE) as “The Environmental Health Action Program” [grant number 2018001360004] and the National Research Foundation of Korea (NRF) grant funded by the Korean government (MSIT) [grant number NRF-2020R1F1A1069853].

Author contributions

Thao Thi Ho, Sanghun Choi: Conceptualization, Methodology, Software, Validation. Thao Thi Ho, Sanghun Choi, Woo Jin Kim, Chang Hyun Lee, Gong Yong Jin, Kum Ju Chae: Resources, Data curation, Writing- Original draft preparation, Writing - Review & Editing. Sanghun Choi: Supervision.

Declaration of competing interest

The authors declare that they have no known competing financial

interests or personal relationships that could have appeared to influence the work reported in this paper.

References

- [1] Y. Lei, Y. Fu, T. Wang, Y. Liu, P. Patel, W.J. Curran, T. Liu, X. Yang, 4D-CT deformable image registration using multiscale unsupervised deep learning, *Phys. Med. Biol.* 65 (2020), 085003.
- [2] K.K. Brock, S. Mutic, T.R. McNutt, H. Li, M.L. Kessler, Use of image registration and fusion algorithms and techniques in radiotherapy: report of the AAPM radiation therapy committee task group No. 132, *Med. Phys.* 44 (2017) e43–e76.
- [3] S. Oh, S. Kim, Deformable image registration in radiation therapy, *Radiat. Oncol. J* 35 (2017) 101–111.
- [4] G.E. Christensen, J.H. Song, W. Lu, I. El Naqa, D.A. Low, Tracking lung tissue motion and expansion/compression with inverse consistent image registration and spirometry, *Med. Phys.* 34 (2007) 2155–2163.
- [5] F.E.-Z.A. El-Gamal, M. Elmogy, A. Atwan, Current trends in medical image registration and fusion, *Egypt. Inf. J.* 17 (2016) 99–124.
- [6] Y. Ou, S.P. Weinstein, E.F. Conant, S. Englander, X. Da, B. Gaonkar, M.K. Hsieh, M. Rosen, A. DeMichele, C. Davatzikos, D. Kontos, Deformable registration for quantifying longitudinal tumor changes during neoadjuvant chemotherapy, *Magn. Reson. Med.* 73 (2015) 2343–2356.
- [7] M. Tan, Z. Li, Y. Qiu, S.D. McMeekin, T.C. Thai, K. Ding, K.N. Moore, H. Liu, B. Zheng, A new approach to evaluate drug treatment response of ovarian cancer patients based on deformable image registration, *IEEE Trans. Med. Imag.* 35 (2016) 316–325.
- [8] Y. Yin, E.A. Hoffman, C.L. Lin, Mass preserving nonrigid registration of CT lung images using cubic B-spline, *Med. Phys.* 36 (2009) 4213–4222.
- [9] J. Ashburner, A fast diffeomorphic image registration algorithm, *Neuroimage* 38 (2007) 95–113.
- [10] B.B. Avants, C.L. Epstein, M. Grossman, J.C. Gee, Symmetric diffeomorphic image registration with cross-correlation: evaluating automated labeling of elderly and neurodegenerative brain, *Med. Image Anal.* 12 (2008) 26–41.
- [11] M.F. Beg, M.I. Miller, A. Trouvé, L. Younes, Computing large deformation metric mappings via geodesic flows of diffeomorphisms, *Int. J. Comput. Vis.* 61 (2005) 139–157.
- [12] X. Yang, R. Kwitt, M. Niethammer, *Fast Predictive Image Registration*, Springer International Publishing, Cham, 2016, pp. 48–57.
- [13] H. Sokooti, B. de Vos, F. Berendsen, B.P.F. Lelieveldt, I. Isgum, M. Staring, *Nonrigid Image Registration Using Multi-Scale 3D Convolutional Neural Networks*, Springer International Publishing, Cham, 2017, pp. 232–239.
- [14] J. Max, S. Karen, Z. Andrew, K. Koray, Spatial Transformer Networks, *NIPS’15: Proceedings of the 28th International Conference on Neural Information Processing Systems*, MIT Press, Montreal, Canada, 2015, pp. 2017–2025.
- [15] H. Sokooti, B. de Vos, F. Berendsen, M. Ghafoorian, S. Yousefi, P.F.B. Lelieveldt, I. Isgum, M. Staring, 3D Convolutional Neural Networks Image Registration Based on Efficient Supervised Learning from Artificial Deformations, *ArXiv*, 2019.
- [16] K.A.J. Eppenhof, J.P.W. Pluim, Pulmonary CT registration through supervised learning with convolutional neural networks, *IEEE Trans. Med. Imag.* 38 (2019) 1097–1105.
- [17] R. Olaf, F. Philipp, B. Thomas, U-net: Convolutional Networks for Biomedical Image Segmentation, *MICCAI*, 2015.
- [18] H. Li, Y. Fan, Non-rigid Image Registration Using Fully Convolutional Networks with Deep Self-Supervision, *Abs/1709*, 2017, 00799.
- [19] G. Wu, M. Kim, Q. Wang, B.C. Munsell, D. Shen, Scalable high-performance image registration framework by unsupervised deep feature representations learning, *IEEE Trans. Biomed. Eng.* 63 (2016) 1505–1516.
- [20] S. Shan, X. Guo, W. Yan, E.I.-C. Chang, Y. Fan, Y.J.A. Xu, Unsupervised End-To-End Learning for Deformable Medical Image Registration, 2017, 08608 abs/1711.
- [21] D. Wang, Y. Pan, O.C. Durumeric, J.M. Reinhardt, E.A. Hoffman, J.D. Schroeder, G. E. Christensen, PLOSS: population learning followed by one shot learning pulmonary image registration using tissue volume preserving and vesselness constraints, *Med. Image Anal.* 79 (2022), 102434.
- [22] B.D. de Vos, F.F. Berendsen, M.A. Viergever, H. Sokooti, M. Staring, I. Isgum, A deep learning framework for unsupervised affine and deformable image registration, *Med. Image Anal.* 52 (2019) 128–143.
- [23] T. Fechter, D. Baltas, One-Shot learning for deformable medical image registration and periodic motion tracking, *IEEE Trans. Med. Imag.* 39 (2020) 2506–2517.
- [24] G. Balakrishnan, A. Zhao, M.R. Sabuncu, J. Guttag, A.V. Dalca, VoxelMorph: a learning framework for deformable medical image registration, *IEEE Trans. Med. Imag.* 38 (2019) 1788–1800.
- [25] Y. Zheng, S. Jiang, Z. Yang, Deformable registration of chest CT images using a 3D convolutional neural network based on unsupervised learning, *J. Appl. Clin. Med. Phys.* 22 (2021) 22–35.
- [26] Y. Fu, Y. Lei, T. Wang, K. Higgins, J.D. Bradley, W.J. Curran, T. Liu, X. Yang, LungRegNet: an unsupervised deformable image registration method for 4D-CT lung, *Med. Phys.* 47 (2020) 1763–1774.
- [27] Y. Hu, M. Modat, E. Gibson, W. Li, N. Ghavami, E. Bonmati, G. Wang, S. Bandula, C.M. Moore, M. Emberton, S. Ourselin, J.A. Noble, D.C. Barratt, T. Vercauteren, Weakly-supervised convolutional neural networks for multimodal image registration, *Med. Image Anal.* 49 (2018) 1–13.
- [28] Z. Jiang, F.F. Yin, Y. Ge, L. Ren, A multi-scale framework with unsupervised joint training of convolutional neural networks for pulmonary deformable image registration, *Phys. Med. Biol.* 65 (2020), 015011.

- [29] J. Yang, J. Yang, F. Zhao, W. Zhang, An unsupervised multi-scale framework with attention-based network (MANet) for lung 4D-CT registration, *Phys. Med. Biol.* 66 (2021).
- [30] S. Zhao, Y. Dong, E. Chang, Y. Xu, Recursive Cascaded Networks for Unsupervised Medical Image Registration, 2019, IEEE/CVF International Conference on Computer Vision (ICCV), 2019, pp. 10599–10609.
- [31] S. Zhao, T. Lau, J. Luo, E.I. Chang, Y. Xu, Unsupervised 3D end-to-end medical image registration with volume tweening network, *IEEE J. Biomed. Health Inf.* 24 (2020) 1394–1404.
- [32] T.C.W. Mok, A.C.S. Chung, Large Deformation Diffeomorphic Image Registration with Laplacian Pyramid Networks, Springer International Publishing, Cham, 2020, pp. 211–221.
- [33] M. Kang, X. Hu, W. Huang, M.R. Scott, M. Reyes, Dual-stream pyramid registration network, *Med. Image Anal.* 78 (2022), 102379.
- [34] J. Ruhaak, T. Polzin, S. Heldmann, I.J.A. Simpson, H. Handels, J. Modersitzki, M.P. Heinrich, Estimation of large motion in lung CT by integrating regularized keypoint correspondences into dense deformable registration, *IEEE Trans. Med. Imag.* 36 (2017) 1746–1757.
- [35] G.E. Christensen, H.J. Johnson, Consistent image registration, *IEEE Trans. Med. Imag.* 20 (2001) 568–582.
- [36] T. Rohlffing, C.R. Maurer Jr., D.A. Bluemke, M.A. Jacobs, Volume-preserving nonrigid registration of MR breast images using free-form deformation with an incompressibility constraint, *IEEE Trans. Med. Imag.* 22 (2003) 730–741.
- [37] R. Castillo, E. Castillo, R. Guerra, V.E. Johnson, T. McPhail, A.K. Garg, T. Guerrero, A framework for evaluation of deformable image registration spatial accuracy using large landmark point sets, *Phys. Med. Biol.* 54 (2009) 1849–1870.
- [38] D. Kuang, Cycle-consistent training for reducing negative jacobian determinant in deep registration networks, in: N. Burgos, A. Gooya, D. Svoboda (Eds.), *Simulation and Synthesis in Medical Imaging*, Springer International Publishing, 2019, pp. 120–129.
- [39] N. Ibtehaz, M.S. Rahman, MultiResUNet : rethinking the U-Net architecture for multimodal biomedical image segmentation, *Neural Network.* 121 (2020) 74–87.
- [40] T.C.W. Mok, A.C.S. Chung, Fast symmetric diffeomorphic image registration with convolutional neural networks, in: Conference on Computer Vision and Pattern Recognition, 2020, 2020.
- [41] H.B. Cho, K.J. Chae, G.Y. Jin, J. Choi, C.L. Lin, E.A. Hoffman, S.E. Wenzel, M. Castro, S.B. Fain, N.N. Jarjour, M.L. Schiebler, R.G. Barr, N. Hansel, C. B. Cooper, E.C. Kleerup, M.K. Han, P.G. Woodruff, R.E. Kanner, E.R. Bleecker, S. P. Peters, W.C. Moore, C.H. Lee, S. Choi, L. National Heart, S. Blood Institute's, C. S., InteRmediate Outcome Measures In, P. Severe Asthma Research, Structural and functional features on quantitative chest computed tomography in the Korean asian versus the white American healthy non-smokers, *Korean J. Radiol.* 20 (2019) 1236–1245.
- [42] T.T. Ho, T. Kim, W.J. Kim, C.H. Lee, K.J. Chae, S.H. Bak, S.O. Kwon, G.Y. Jin, E. K. Park, S. Choi, A 3D-CNN model with CT-based parametric response mapping for classifying COPD subjects, *Sci. Rep.* 11 (2021) 34.
- [43] E. Castillo, R. Castillo, J. Martinez, M. Shenoy, T. Guerrero, Four-dimensional deformable image registration using trajectory modeling, *Phys. Med. Biol.* 55 (2010) 305–327.
- [44] R. Castillo, E. Castillo, D. Fuentes, M. Ahmad, A.M. Wood, M.S. Ludwig, T. Guerrero, A reference dataset for deformable image registration spatial accuracy evaluation using the COPDGene study archive, *Phys. Med. Biol.* 58 (2013) 2861–2877.
- [45] K. Krishnan, L. Ibanez, W.D. Turner, J. Jomier, R.S. Avila, An open-source toolkit for the volumetric measurement of CT lung lesions, *Opt Express* 18 (2010) 15256–15266.
- [46] X. H, J. G, X. Z, H. B, S. G, D. K, A. M, E. H, J. R, R.G. B, E. A, A. L, Recursive refinement network for deformable lung registration between exhale and inhale CT scans, *ArXiv* (2021), 07608, 2016.
- [47] S. Choi, E.A. Hoffman, S.E. Wenzel, M.H. Tawhai, Y. Yin, M. Castro, C.L. Lin, Registration-based assessment of regional lung function via volumetric CT images of normal subjects vs. severe asthmatics, *J. Appl. Physiol.* 115 (2013) 730–742, 1985.
- [48] M. Modat, J.R. McClelland, Lung Registration Using the NiftyReg Package, 2010.
- [49] J. Zhang, Inverse-Consistent Deep Networks for Unsupervised Deformable Image Registration, 2018, 03443. ArXiv, abs/1809.
- [50] J. Long, E. Shelhamer, T. Darrell, Fully convolutional networks for semantic segmentation, in: 2015 IEEE Conference on Computer Vision and Pattern Recognition, CVPR, 2015, pp. 3431–3440.