J. Shanghai Jiao Tong Univ. (Sci.), 2022
https://doi.org/10.1007/s12204-022-2525-3

Dlung: Unsupervised Few-Shot Diffeomorphic Respiratory Motion Modeling

CHEN Peizhi[1,2*] (陈培芝), *GUO Yifan*[1] (郭逸凡), *WANG Dahan*[1,2] (王大寒), *CHEN Chinling*[1,3,4*] (陈金铃)

(1. College of Computer and Information Engineering, Xiamen University of Technology, Xiamen 361024, Fujian, China; 2. Fujian Key Laboratory of Pattern Recognition and Image Understanding, Xiamen 361024, Fujian, China; 3. School of Information Engineering, Changchun Sci-Tech University, Changchun 130600, China; 4. Department of Computer Science and Information Engineering, Chaoyang University of Technology, Taichung 41349, Taiwan, China)

Abstract: Lung image registration plays an important role in lung analysis applications, such as respiratory motion modeling. Unsupervised learning-based image registration methods that can compute the deformation without the requirement of supervision attract much attention. However, it is noteworthy that they have two drawbacks: they do not handle the problem of limited data and do not guarantee diffeomorphic (topology-preserving) properties, especially when large deformation exists in lung scans. In this paper, we present an unsupervised few-shot learning-based diffeomorphic lung image registration, namely Dlung. We employ fine-tuning techniques to solve the problem of limited data and apply the scaling and squaring method to accomplish the diffeomorphic registration. Furthermore, atlas-based registration on spatio-temporal (4D) images is performed and thoroughly compared with baseline methods. Dlung achieves the highest accuracy with diffeomorphic properties. It constructs accurate and fast respiratory motion models with limited data. This research extends our knowledge of respiratory motion modeling.

Key words: unsupervised few-shot learning, respiratory motion modeling, diffeomorphic registration
CLC number: TP 751, R 816.41 **Document code:** A

0 Introduction

Lung image registration is the process of constructing a dense correspondence between lung image pairs. It is critical for the analysis of the lungs, for example, respiratory motion modeling. Respiratory motion modeling builds a dense spatial correspondence using spatio-temporal (4D) images to describe respiratory motion and aid the analysis of thoracic organs such as the lungs. According to the World Health Organization report[1], lung cancer is the leading cause of death. Radiotherapy is one of the main treatments for lung cancer patients (i.e., operation, chemotherapy, and radiotherapy). Respiratory motion modeling can help radiotherapy correctly track cancer and avoid damage to normal tissues, thus boosting the radiotherapy effect.

Traditional lung image registration methods such as HAMMER[2] and Elastix[3] iteratively optimize an objective function to obtain the transformation. They often require significant amounts of computation time (ranging from tens of minutes to several hours) using central processing units (CPUs), and it is difficult to accelerate iterative algorithms using graphic processing units (GPUs). Among the traditional methods, diffeomorphic registrations such as large displacement diffeomorphic metric mapping (LDDMM)[4], diffeomorphic anatomical registration through exponentiated lie algebra (DARTEL)[5], and standard symmetric normalization (SyN)[6] have become well-established methods. While such traditional methods can achieve accurate results, they suffer from high computational costs, which impedes their use in time-sensitive applications such as image-guided surgery and radiation treatment.

Deep learning-based methods can overcome computationally expensive problems by using GPUs. Furthermore, deep learning-based methods consistently outperform traditional optimization-based techniques[7]. There are two main kinds of algorithms: supervised and unsupervised methods. Supervised methods such as RegNet[8], Eppenhof and Pluim's method[9] learn the displacement field using the ground-truth field. However, the acquirement of ground truth displacement

Received: 2021-01-27 **Accepted:** 2021-09-17
Foundation item: the National Natural Science Foundation of China (No. 61801413), and the Natural Science Foundation of Fujian Province (Nos. 2019J05123 and 2017J05110)
∗**E-mail:** pzc@xmut.edu.cn, 2018000025@xmut.edu.cn

can be a difficult task. Some methods use a synthetic displacement field that cannot represent a real transformation[8], while other methods use the transformation learned from existing traditional methods[9] which are limited by the traditional method and require significant training time.

Compared with supervised methods, unsupervised methods can learn the displacement field from the images directly without the need for ground-truth annotation. Unsupervised learning-based image registration has therefore gained huge interest in this field of research[10]. A common method used is VoxelMorph[11], which learns the displacement field through a U-Net framework by unsupervised learning. VoxelMorph can achieve similar accuracy at a greater order of magnitude faster than traditional methods, but the displacement fields do not have topological-preserving properties.

Currently, most work in this field focuses on fast and accurate image registration via unsupervised learning models that learn the deformation[11-19]. Such models can achieve fast and accurate registration results on different body parts, for example, brain[11,17], liver[17], thoracic region[14,16], and abdominal region[13]. However, image registration with large deformation in the lungs requires more topology-preserving results. In other words, they do not consider topology-preservation and therefore cannot represent the large deformation nature. Other kinds of methods that focus on diffeomorphic registration also exist[18-22] whilst the focus has been on brain[18-19,21-22] and cardiac regions[20]. Hu et al.[23] investigated a diffeomorphic registration on lung image registration while its application to respiratory motion modeling is still unknown.

Moreover, the respiratory dataset is limited for deep learning-based methods due to the high cost of spatio-temporal image acquisition and the privacy of medical images. Unsupervised few-shot learning can transfer knowledge from the source domain to the target domain in an unsupervised manner. Unsupervised few-shot learning has been recently applied in medical image analysis such as classification[24], segmentation[25-26], and registration[27]. Ferrante et al.[27] showed the adaptability of unsupervised convolutional neural networks (CNN)-based deformable image registration to unseen image domains. Besides, there are some unsupervised-based respiratory motion modeling investigations[9,12,28]. However, they ignored diffeomorphic registration.

To summarize, the most current techniques in the lung image registration field possess the following drawbacks: ① they do not provide topological-preserving properties, especially when large deformation exists in the lungs, and ② there is no focus on the problem of limited data.

In the paper, an unsupervised few-shot learning method for diffeomorphic lung image registration is introduced, named Dlung. Inspired by the previous work[18], we address the topological-preserving properties by diffeomorphic registration and adopt few-shot learning to handle the issue of limited data.

Experiments on spatio-temporal computed tomography (4D-CT) images from 4D-Lung[29-32] are conducted and analyzed in terms of accuracy and runtime. Compared with SyN, Elastix, and VoxelMorph, Dlung is more accurate and faster. Here, we also provide an application for respiratory motion modeling, which can aid radiologists in the delineation of clinical target volumes and interactive image registration. This research extends our knowledge of respiratory motion modeling.

Overall, the contributions of the work include the following.

(1) We examine the diffeomorphic registration on lung scans.

(2) We address the problem of limited data in an unsupervised few-shot learning-based image registration task.

1 Method

The aim of this paper is to construct an unsupervised few-shot learning-based spatio-temporal registration method with diffeomorphic properties and apply it in respiratory motion modeling. There are three procedures: ① pre-processing; ② unsupervised few-shot learning-based image registration between two volumes; ③ spatio-temporal image registration. Of these, the 2nd step forms the core function that will be explained in detail later, while the 1st and 3rd steps are explained in the implementation subsection.

To handle the registration between two images, we propose an unsupervised learning-based image registration model, as shown in Fig. 1. First, a U-Net-like network appraises displacement flow between two images. Then, the scaling and squaring (SS) layer computes the diffeomorphic registration whose displacement field with topology-preserving properties is obtained. Finally, through spatial interpolation, the moved image is obtained. A volume from the end-exhale phase is selected as the atlas so that other volumes can be registered to the atlas. The loss function is evaluated on similarity and smoothness.

The training process is represented by the dashed lines while the testing process is the solid lines. Given a fixed image $\boldsymbol{I}_{\mathrm{F}}$ and a moving image $\boldsymbol{I}_{\mathrm{M}}$, the geometric transformation that can align $\boldsymbol{I}_{\mathrm{M}}$ to $\boldsymbol{I}_{\mathrm{F}}$ is computed. The model is designed to realize fast diffeomorphic image registration consisting of four sections.

U-Net-like Network The displacement flow of the pair $(\boldsymbol{I}_{\mathrm{F}}, \boldsymbol{I}_{\mathrm{M}})$ is calculated using a U-Net-like network.

SS Layer The flow is further improved by adding an SS layer to provide diffeomorphic properties.

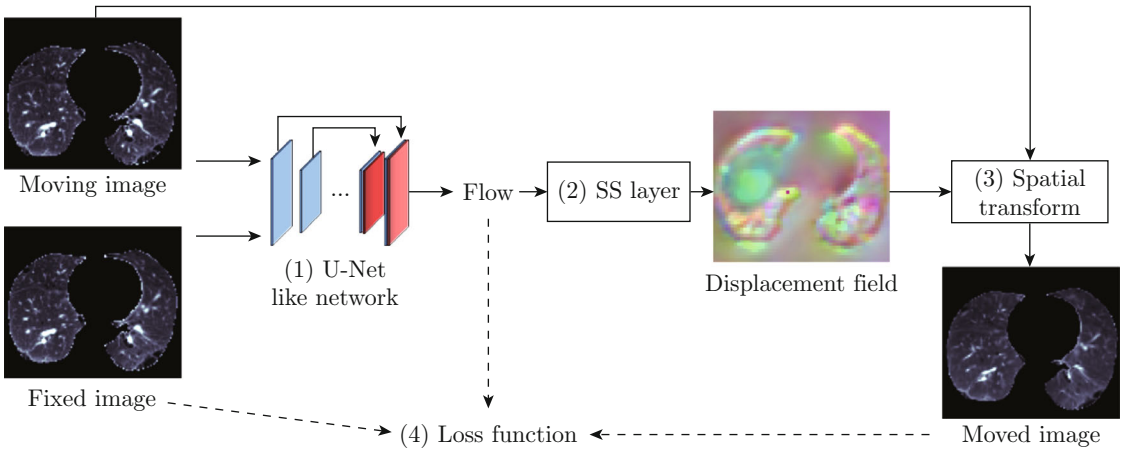


Fig. 1 Overview of the image registration method on lung images based on unsupervised learning

Spatial Transform A spatial transform layer (STL) is applied on $\boldsymbol{I}_{\mathrm{M}}$ to achieve the moved image $\boldsymbol{I}'_{\mathrm{M}}$.

Loss Function The training is carried on by minimizing the loss function.

These four sections are discussed in more detail below.

1.1 U-Net-like Network

U-Net-like networks are applied to generate a coarse displacement field between $\boldsymbol{I}_{\mathrm{F}}$ and $\boldsymbol{I}_{\mathrm{M}}$. The network framework adopted in this work is shown in Fig. 2. The network is an encoder-decoder structure with skip connections. The number in each layer represents its channel number while the number below represents the scaling factor with the fixed and moving images.

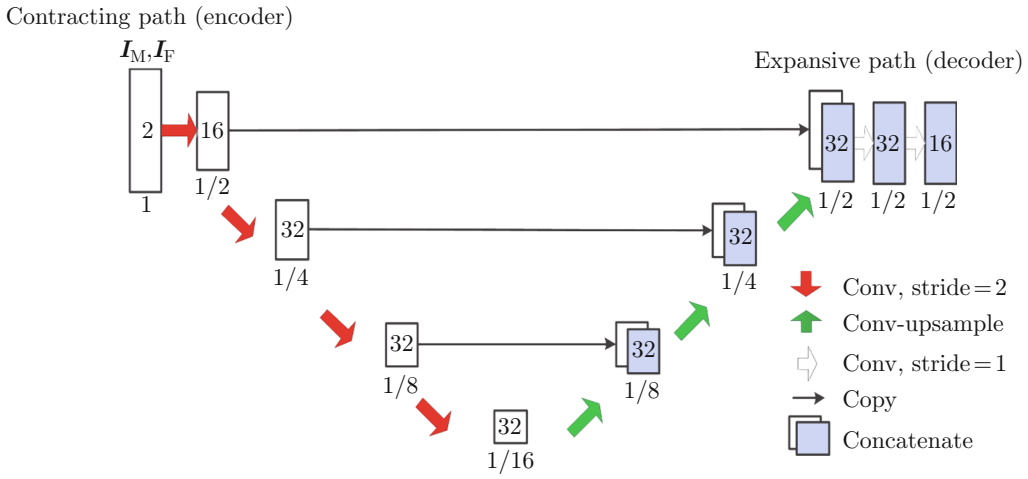


Fig. 2 U-Net structure

The U-Net-like network can be divided into two parts: the contracting path and the expansive path. The contracting path does extracts in a down-sampling step-by-step procedure and plays a role in feature extraction on different scales, while the expansive path does extracts in an up-sampling step-by-step procedure and ensures the output's size. During the expansive path, skip connections are employed to enrich the feature from the contracting path, i.e., the corresponding features (from images of the same size) in the contracting path are concatenated to the expansive path.

In the convolution layers, $3 \times 3 \times 3$ convolutions with Gaussian kernels are used. A stride of 2 is used in the contracting path to down-sample the images and nearest-neighbor interpolation is used in the expansive path to up-sample the images.

1.2 SS Layer

The U-Net-like network can provide the coarse registration without the diffeomorphic property, so it cannot properly represent the true deformation. In this paper, the SS layer[18] is utilized to yield the diffeomorphic property, briefly reviewed in the following section and shown in Fig. 3. In the initialization step, we initialize the stationary velocity field. Since the integration of the stationary velocity field represents a one-parameter subgroup of diffeomorphism, the exponential field of the

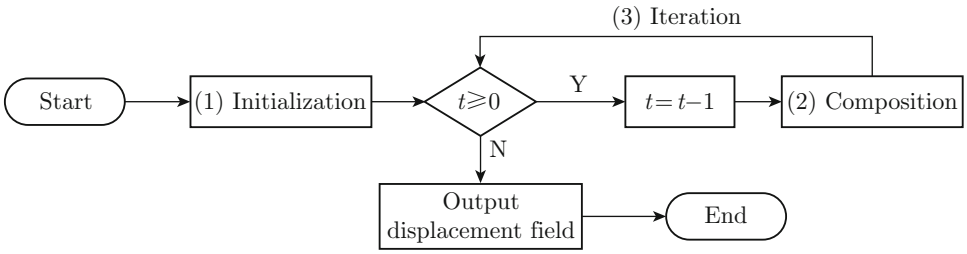


Fig. 3 Flowchart of the SS layer

stationary velocity field is computed. Then, in the composition step, the displacement field is further interpolated. The composition step is iterated to achieve the final displacement field.

The SS layer stems from the theory of diffeomorphic registration whose transformations (and their inverse) are one-to-one and regular. There are different ways to implement diffeomorphic registration. Let the geometric transformation between two images be $\boldsymbol{\Phi} : \mathbf{R}^3 \to \mathbf{R}^3$. We calculate $\boldsymbol{\Phi}$ through the following ordinary differential equation (ODE):

$$\frac{\partial \boldsymbol{\Phi}^{(t)}}{\partial t} = \nu(\boldsymbol{\Phi}^{(t)}), \tag{1}$$

where $\nu(\boldsymbol{\Phi}^{(t)})$ is the stationary velocity of $\boldsymbol{\Phi}$ at time t and its initial value is $\nu(\boldsymbol{\Phi}^0)$. The target transformation is the integration of velocity fields over $t \in [0, 1]$, that is $\boldsymbol{\Phi}^{(1)} = \int_0^1 \nu(\boldsymbol{\Phi}^{(t)}) \mathrm{d}t$.

To achieve the deformation, the SS layer computes the integration of the stationary velocity field in Eq. (1) numerically. The integration represents a one-parameter subgroup of diffeomorphisms. In group theory, ν is a member of the Lie algebra: ν is exponentiated to produce $\boldsymbol{\Phi}(t) = e^{\nu}$ so that $\boldsymbol{\Phi}(t)$ is a member of the Lie group. Subsequently, $\boldsymbol{\Phi}\Big(\frac{1}{2^{t-1}}\Big) = \boldsymbol{\Phi}\Big(\frac{1}{2^t}\Big) \circ \boldsymbol{\Phi}\Big(\frac{1}{2^t}\Big)$, where $t = T, T-1, \cdots, 1, 0$, and $\circ$ is composition map with the Lie group associated. We can iterate t from T to 0 to achieve $\boldsymbol{\Phi}(1)$. There are three procedures.

Initialization We initialize $t = T$ and $\boldsymbol{\Phi}\Big(\frac{1}{2^t}\Big) = \boldsymbol{\Phi}(0) + z$, where z is sampled from the displacement flow. We set $T = 7$ so that $\boldsymbol{\Phi}(0) \sim \boldsymbol{\Phi}\Big(\frac{1}{2^T}\Big)$. $z \sim N(\boldsymbol{\Phi}(0); \mu, \Sigma)$, where μ is the mean and Σ is the variance of $\boldsymbol{\Phi}(0)$.

Composition We interpolate $\boldsymbol{\Phi}\Big(\frac{1}{2^t}\Big)$ to attain $\boldsymbol{\Phi}\Big(\frac{1}{2^{t-1}}\Big)$.

Iteration We iterate from $t = T$ to $t = 0$ to attain $\boldsymbol{\Phi}\Big(\frac{1}{2^0}\Big) = \boldsymbol{\Phi}(1)$.

Importantly, the SS method computes the displacement field through the iteration of diffeomorphisms, after which it remains as diffeomorphism since it agrees with group theory. Here, we can obtain the displacement field with diffeomorphic properties guaranteed.

1.3 Spatial Transform

After the refinement of the SS layers, the displacement filed $\boldsymbol{\Phi}$ with diffeomorphic properties can be attained and used to transform the moving image $\boldsymbol{I}_{\mathrm{M}}$ into a moved image $\boldsymbol{I}'_{\mathrm{M}}$, i.e., $\boldsymbol{I}'_{\mathrm{M}}(x) = \boldsymbol{I}_{\mathrm{M}}(\boldsymbol{\Phi}(x))$. To ensure the usability of standard gradient-based optimization, the spatial transformation layer[7] is applied to make the transformation differentiable.

1.4 Loss Function

In the unsupervised learning-based model, it is desired that the transformation minimizes the difference between the moved and fixed images and thus satisfies the smoothness regulation of the displacement field, which can be a man-made supervised signal. Thus, the definition of the loss function plays an unsupervised role in image registration.

Here, we define the similarity of the moved and fixed images as

$$E_{\mathrm{S}} = \frac{1}{|\Omega|} \sum_{x \in \Omega} (\boldsymbol{I}_{\mathrm{M}}(\boldsymbol{\Phi}(x)) - \boldsymbol{I}_{\mathrm{F}}(x))^2, \tag{2}$$

where $\boldsymbol{\Phi}(x)$ is the transformed x in the 3D spatial domain Ω.

To obtain a smooth deformation, the regulation of the displacement field is given:

$$E_{\mathrm{R}} = \frac{1}{|\Omega|} \sum_{x \in \Omega} d(x)^2, \tag{3}$$

where $d(x)$ is the derivative of x.

The total loss function is then given:

$$E = E_{\mathrm{S}} + \lambda E_{\mathrm{R}}, \tag{4}$$

where λ is treated as a hyperparameter.

1.5 Few-Shot Learning

Due to the high acquisition cost, only a limited number of 4D-CT images could be used for respiratory motion modeling, leading it to an unsupervised few-shot learning problem. A fine-tuning method is therefore implemented to solve this issue. Here, we fine-tune the pre-trained model from VoxelMorph[18], which is an atlas-based image registration method trained for brain images.

1.6 Implementation

1.6.1 *Pre-Processing*

Our pre-processing procedures include the following steps, as shown in Fig. 4.

Resample The thoracic CT images were resampled to isotropic.

Lung Segmentation The existing research[29] shows that respiratory motion modeling on lung region of interest (ROI) performs more accurately than the whole chest CT images. Hence, we conduct respiratory motion modeling on lung ROI. Herein, -500 HU is used as the threshold to segment the images and morphological processing to achieve a complete connected lung region.

Image Cropping Lung images are cropped to the same resolution of $160 \times 192 \times 224$.

Since the region with a range of $[-1\,000, 400]$ HU is of most interest, pixels are normalized to $[0, 1]$ after pre-processing steps.

(a) Resample to isotropic (b) Lung segmentation (c) Image cropping

Fig. 4 Pre-processing diagram

1.6.2 *Spatio-Temporal Image Registration*

In the spatio-temporal image registration application, the input is spatio-temporal images whose volumes are from different phases of a whole respiratory period. To construct the respiratory motion model, the volumes in the spatio-temporal images need to be registered to a common space. To do this, an atlas-based registration scheme is adopted, i.e., all the volumes are registered to the same atlas. In this paper, the volume from the end of expiration is chosen to be the atlas.

2 Experimental Set-Up and Results

In this paper, we propose Dlung for the lung image registration task which is the basic of respiratory motion modeling, whose input is spatio-temporal CT images from different respiratory phases, while the output is the displacement field. We adopt atlas-based registration, i.e., we select a volume as the atlas and register other volumes to it. In the training procedure, we choose a common volume of the end of expiratory as the atlas. While in the testing procedure, we use a specified atlas that is from the end of expiratory of that spatio-temporal image.

2.1 Materials

In this study, a 4D-Lung[29-32] dataset constructed from 20 patients with non-small cell lung cancers was used. All 4D-Lung images were collected using the same device with the same settings. The 4D-Lung image slice thicknesses were 3 mm. The slice number of the volume was ranging from dozens to more than one hundred. Since data diversity without enough training data will lead to overfitting, we selected images whose slice number was similar in our experiments. Specifically, we chose 4 pairs of spatio-temporal images for training and 5 for testing. Dlung was implemented using Python3 and Keras with a Tensorflow backend. A NVIDIA 2080ti GPU and an Intel Xeon E5-2678v3 CPU were used.

2.2 Evaluation Measurement

Dice[33] Dice evaluates the overall registration accuracy on segmentation and plays the role of coarse accuracy evaluation given by

$$\mathrm{Dice}(S_{\mathrm{M}(\boldsymbol{\Phi})}, S_{\mathrm{F}}) = 2\frac{S_{\mathrm{M}(\boldsymbol{\Phi})} \cap S_{\mathrm{F}}}{S_{\mathrm{M}(\boldsymbol{\Phi})} + S_{\mathrm{F}}}, \tag{5}$$

where S_{F} and $S_{\mathrm{M}(\boldsymbol{\Phi})}$ are the segmentation of the fixed and moved images, respectively.

Jacobian Determinant[18] The Jacobian determinant of each pixel captures the local properties of its deformation, which is

$$|\boldsymbol{J}| = \begin{vmatrix} \dfrac{\partial y_1}{\partial x_1} & \dfrac{\partial y_1}{\partial x_2} & \dfrac{\partial y_1}{\partial x_3} \\ \dfrac{\partial y_2}{\partial x_1} & \dfrac{\partial y_2}{\partial x_2} & \dfrac{\partial y_3}{\partial x_3} \\ \dfrac{\partial y_3}{\partial x_1} & \dfrac{\partial y_3}{\partial x_2} & \dfrac{\partial y_3}{\partial x_3} \end{vmatrix}, \tag{6}$$

where, x and y are the locations of the fixed and moved images respectively; $|\cdot|$ is the determinant of a matrix. The Jacobian determinant is calculated with discrete

gradients on each location of the displacement field. The negative number of $|\boldsymbol{J}|$ represents the pixels with folding in the displacement field.

2.3 Baseline Methods

VoxelMorph and SyN are state-of-the-art unsupervised learning-based and traditional optimization-based algorithms while Elastix is widely applied in lung image registration. To validate the accuracy and efficiency of Dlung, comparisons with VoxelMorph[18], SyN[6], and Elastix[3] were conducted. We applied the source code provided by VoxelMorph①, ANTsPy② for SyN and Elastix.

2.4 Comparison with State-of-the-Arts

We conduct main experiments on spatio-temporal image datasets by comparing Dlung with 3 state-of-the-art methods: Elastix, SyN, and VoxelMorph.

Table 1 presents the average Dice score calculated on 5 pairs of spatio-temporal images. Both Dice scores and percentages of negative Jacobian determinants are displayed in the form of the mean (variance). The higher the Dice score, the better, and the lower the percentage of $|\boldsymbol{J}| \leqslant 0$, the better. Dlung attains Dice score around 0.015 higher than Elastix and SyN, and 0.01 higher than VoxelMorph. In addition, Table 1 also shows the average percentage of pixels whose Jacobian determinant is below 0 (i.e., $|\boldsymbol{J}| \leqslant 0$). Dlung gains the percentage close to 0. In other words, Dlung can produce the registration result without folding to a large extent. The percentage of possible folding pixels is almost 2 000 times (for VoxelMorph) and 100 times (for Elastix) larger than that of Dlung.

Table 1 Average accuracy comparison

Method	Dice score	$\lvert\boldsymbol{J}\rvert \leqslant 0/\%$
Elastix[3]	0.942 (0.013)	1.37×10^{-2} (2×10^{-2})
SyN[6]	0.939 (0.017)	0
VoxelMorph[18]	0.946 (0.025)	0.30 (0.12)
Dlung	**0.955 (0.013)**	$\mathbf{1.4 \times 10^{-4}}$ ($\mathbf{2.7 \times 10^{-4}}$)

Figure 5 presents the accuracy comparison on different patients. We conducted experiments on 5 pairs of spatio-temporal images and evaluated the Dice scores. Dlung is optimal in four. The box diagram of Dlung is flatter than that of VoxelMorph, indicating Dlung performs more stable than VoxelMorph.

Figure 6 shows accuracy comparison on different respiratory phases. There are 10 volumes that represent 10 respiratory peoriods in each spatio-temporal image. Hence, as the phase becomes larger, the deformation becomes larger during the inspiratory cycle while the deformation becomes smaller during the expiratory cycle. The deformation between the zeroth and fourth phases is the largest. As Fig. 6 shows, when the deformation increases, the Dice score of Dlung remains stable compared with VoxelMorph.

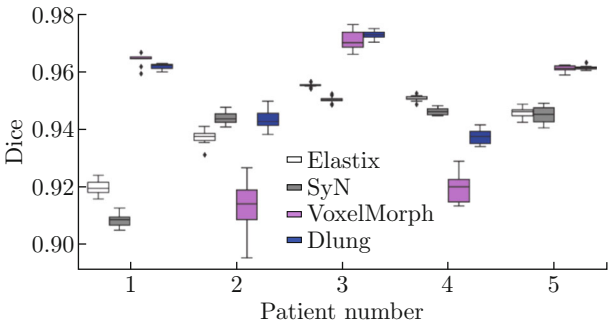


Fig. 5 Accuracy comparison on different patients

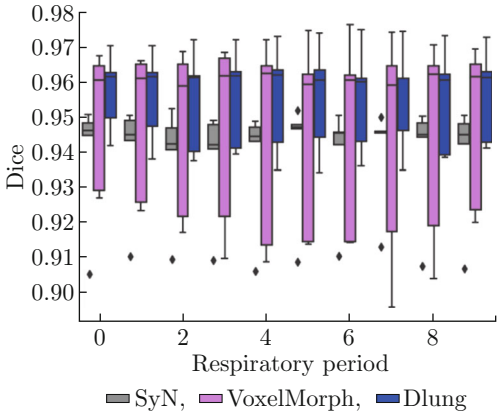


Fig. 6 Accuray comparison on different respiratory phases

Table 2 presents the runtime comparison. Dlung obtains a runtime of two orders of magnitude times faster than SyN, which can be attributed to the GPU. For example, approximately 0.3 s is required to register two images for Dlung, and around $10 \times 0.3\,\mathrm{s} = 3\,\mathrm{s}$ to construct a respiratory motion model. In contrast, the traditional SyN method needs about $10 \times 20\,\mathrm{s} = 200\,\mathrm{s}$ (i.e., more than 3 min) and traditional Elastix method needs about $10 \times 44\,\mathrm{s} = 440\,\mathrm{s}$ (i.e., more than 7 min). Besides, the runtime of Dlung is 0.1 s slower than that of VoxelMorph because Dlung has an extra diffeomorphic operation.

Table 2 Runtime comparison

Method	Runtime/s
Elastix[3]	44.26 (2.67)
SyN[6]	19.73 (1.02)
VoxelMorph[18]	0.20 (0.05)
Dlung	**0.30** (**0.08**)

2.5 Analytical Study

To thoroughly evaluate the proposed Dlung algorithm, a variety of ablation studies were performed,

①https://github.com/voxelmorph/voxelmorph
②https://github.com/ANTsX/ANTsPy

which include diffeomorphic registration and few-shot learning.

2.5.1 *Influence of SS Layer*

The goal of this ablation is to analyze the effectiveness of the SS layer. In the ablation study, we remove the SS layer from Dlung, and investigate its performance. The result is provided in Table 3. The result shows that the SS layer can help Dlung to reduce the percentage of $|\boldsymbol{J}| < 0$ to zero and attain diffeomorphic registration.

Table 3 Ablation study on the SS layer

Method	Dice score	$\lvert\boldsymbol{J}\rvert < 0/\%$
Without SS	0.89 (0.038 5)	1.66 (0.079)
With SS	0.92 (0.035 4)	0

2.5.2 *Influence of Few-Shot Learning*

To demonstrate the effectiveness of few-shot learning, we design an ablation experiment to compare two manners: train from scratch (without few-shot learning) and fine-tuning (with few-shot learning). We trained a model from scratch without pre-trained models and fine-tuned a pre-trained model from brain image registration. We made a comparison on the Dice score. The results were shown in Fig. 7. We can observe that the algorithm with fine-tuning (with_FT) performs better than the one without fine-tuning (without_FT), which shows the effectiveness of fine-tuning in Dlung.

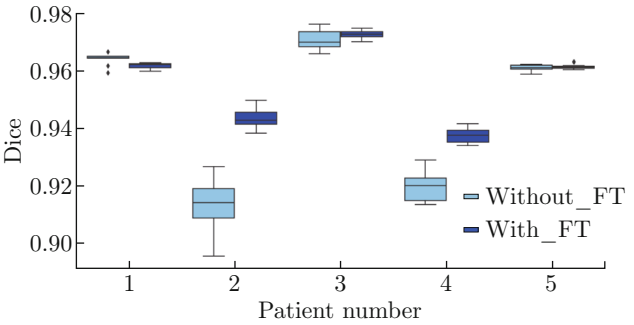
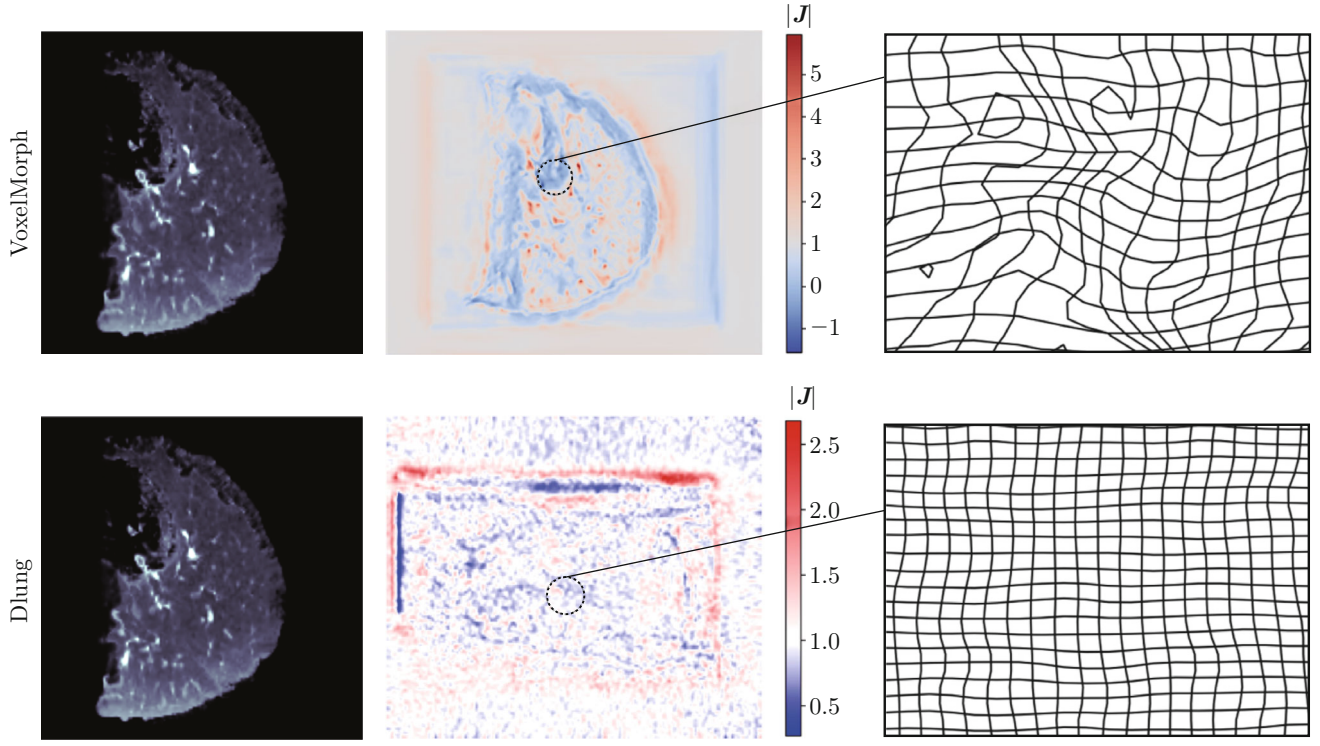


Fig. 7 Ablation study on few-shot learning

2.6 Visualization

2.6.1 *Visualization of Displacement Field*

In addition to the quantitative analysis, we also observe the qualitative analysis. We displayed the Jacobian determinant and the displacement field of VoxelMorph and Dlung. As shown in Fig. 8, Both VoxelMorph and Dlung use the same fixed and moving images. The Jacobian determinant and its warp field in a sub-region demonstrate the effectiveness of the diffeomorphic registration. The heat map represents the $|\boldsymbol{J}|$ value of each pixel. However, Dlung provides a smooth displacement field, while VoxelMorph provides a displacement field with folding.



(a) Moved image (b) Jacobian determinant (c) Display of displacement field

Fig. 8 Visualization of diffeomorphic registration on VoxelMorph and Dlung

2.6.2 *Visualization of Respiratory Motion Modeling*

A respiratory motion model is built using Dlung on a pair of 4D-CT images and the visualization result is shown in Fig. 9. Note that Dlung offers a dense displacement field in solid space while a discrete displacement field on the lung surface is shown.

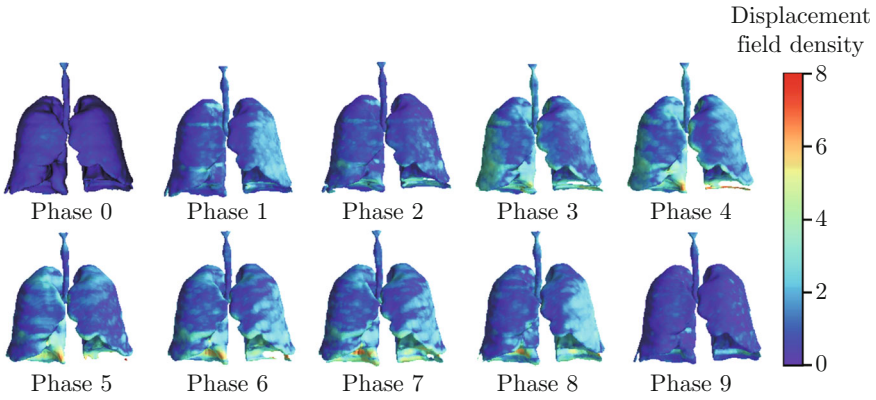


Fig. 9 Application of respiratory motion modeling

3 Discussion

We introduce an unsupervised few-shot learning-based registration method with diffeomorphic properties, named Dlung, and present its application in respiratory motion modeling. The method is constructed through an unsupervised learning method without the need of expensive annotation supervised signals. We employ few-shot learning to solve the limited data problem. Moreover, it provides diffeomorphic registration results to represent the true deformation more accurately.

We compare Dlung with baseline methods on accuracy. Dlung achieves 0.015 promotion on Dice score compared with SyN and Elastix, and 0.01 promotion with VoxelMorph. On the Jacobian determinant metric, Dlung can achieve the least percentage of negative Jacobian determinant pixels (less than 0.000 14%). This is in stark contrast to VoxelMorph which attains a large number of negative values (0.3% of pixels). The least negative Jacobian determinant represents the least folding pixels. In general, Dlung performs more accurately than state-of-the-art methods.

Furthermore, in the runtime comparison, to register volume images, Dlung takes only 0.3 s, while the traditional method Elastix and SyN need 44 s and 20 s respectively, making Dlung more suitable to be applied in real-time applications. It is also worth noting that VoxelMorph needs 0.1 s less time, VoxelMorph cannot provide accurate registration results.

The ablation study results on the spatio-temporal image registration validate that: the few-shot learning results verify the effectiveness of comparing train-from-scratch; the effectiveness of the SS layer is examined in respiratory motion modeling.

In the visualization study, we display the displacement fields of Dlung and VoxelMorph. The deformation fields of Dlung and VoxelMorph are different. The reason is that Dlung introduces the SS layer, which is a global operation so that the SS layer can change the displacement field globally.

The limitations of Dlung are: the public spatio-temporal dataset is limited; the SS layer has a global effect for diffeomorphic registration. To address these limitations, we will perform data augmentation and explore a localized diffeomorphic registration method in the future.

In the end, an application of respiratory motion modeling is provided. Respiratory motion modeling can help radiotherapy instruments to target tumors in the lungs to improve the efficiency of radiotherapy. Diffeomorphic registration can provide a regular and one-to-one mapping to reflect the real deformation between images, i.e., a mapping without folding. In the future, we will investigate automatic target delineation applications to help radiologists to increase efficiency to make radiotherapy plans.

4 Conclusion

In this work, we propose an unsupervised few-shot learning method with diffeomorphic registration for lung image registration, named Dlung, and apply it in respiratory motion modeling. We employ the SS method to provide the diffeomorphic (i.e., topology-preserving) properties and apply fine-tuning on small 4D CT datasets to handle the problem of data unavailability.

In our experiments, Dlung can provide registration results with diffeomorphic properties with the best accuracy on the Dice score that is evaluated on the lung ROI. Moreover, Dlung is able to construct a respiratory

motion model in less than 3 s on a GPU, significantly faster than the traditional SyN method requiring more than 7 min.

The limitations are as follows: first, we focus on the application value of Dlung in lung motion modeling, while the method lacks novelty; second, the lack of public spatio-temporal images limits further experimental verification of robustness; third, the SS layers is a global operation so that it can change the displacement field globally. For the above limitations, we plan to conduct data augmentation to further enhance the robustness and seek localized diffeomorphic registration.

The work can be used as a basic framework of respiratory motion modeling. In the future, Dlung will be applied to some radiotherapy-related tasks such as the delineation of clinical target volumes and interactive image registration.

Acknowledgement We thank Dr. Geoff Hugo to make the 4D-Lung dataset open access. We also thank Dalca from MIT who makes VoxelMorph source code public in GitHub and keeps it well maintained.

References

[1] STEWART B W, WILD C P. World cancer report 2014 [M]. Lyon: International Agency for Research on Cancer, 2014.

[2] SHEN D G, DAVATZIKOS C. HAMMER: Hierarchical attribute matching mechanism for elastic registration [J]. *IEEE Transactions on Medical Imaging*, 2002, **21**(11): 1421-1439.

[3] KLEIN S, STARING M, MURPHY K, et al. Elastix: A toolbox for intensity-based medical image registration [J]. *IEEE Transactions on Medical Imaging*, 2010, **29**(1): 196-205.

[4] BEG M F, MILLER M I, TROUVÉ A, et al. Computing large deformation metric mappings via geodesic flows of diffeomorphisms [J]. *International Journal of Computer Vision*, 2005, **61**(2): 139-157.

[5] ASHBURNER J. A fast diffeomorphic image registration algorithm [J]. *NeuroImage*, 2007, **38**(1): 95-113.

[6] AVANTS B B, EPSTEIN C L, GROSSMAN M, et al. Symmetric diffeomorphic image registration with cross-correlation: Evaluating automated labeling of elderly and neurodegenerative brain [J]. *Medical Image Analysis*, 2008, **12**(1): 26-41.

[7] JADERBERG M, SIMONYAN K, ZISSERMAN A, et al. Spatial transformer networks [M]//Advances in neural information processing systems 28. Red Hook: Curran Associates, 2015: 1-9.

[8] SOKOOTI H, VOS B D, BERENDSEN F, et al. Nonrigid image registration using multi-scale 3D convolutional neural networks [M]//Medical image computing and computer assisted intervention – MICCAI 2017. Cham: Springer, 2017: 232-239.

[9] EPPENHOF K J, LAFARGE M W, MOESKOPS P, et al. Deformable image registration using convolutional neural networks [J]. *Proceedings of SPIE*, 2018, **10574**: 105740S.

[10] NAZIB A, FOOKES C, PERRIN D. A comparative analysis of registration tools: Traditional vs deep learning approach on high resolution tissue cleared data [DB/OL]. (2018-10-18). https://arxiv.org/abs/1810.08315.

[11] BALAKRISHNAN G, ZHAO A, SABUNCU M R, et al. VoxelMorph: A learning framework for deformable medical image registration [J]. *IEEE Transactions on Medical Imaging*, 2019, **38**(8): 2019Feb4.

[12] FU Y B, LEI Y, WANG T H, et al. LungRegNet: An unsupervised deformable image registration method for 4D-CT lung [J]. *Medical Physics*, 2020, **47**(4): 1763-1774.

[13] HEINRICH M P, HANSEN L. Highly accurate and memory efficient unsupervised learning-based discrete CT registration using 2.5D displacement search [M]//Medical image computing and computer assisted intervention – MICCAI 2020. Cham: Springer, 2020: 190-200.

[14] HERING A, HELDMANN S. Unsupervised learning for large motion thoracic CT follow-up registration [J]. *Proceedings of SPIE*, 2019, **10949**: 109491B.

[15] LEE M C H, OKTAY O, SCHUH A, et al. Image-and-spatial transformer networks for structure-guided image registration [M]//Medical image computing and computer assisted intervention — MICCAI 2019. Cham: Springer, 2019: 337-345.

[16] SENTKER T, MADESTA F, WERNER R. GDL-FIRE^{4D}: Deep learning-based fast 4D CT image registration [M]//Medical image computing and computer assisted intervention — MICCAI 2018. Cham: Springer, 2018: 765-773.

[17] ZHAO S Y, DONG Y, CHANG E, et al. Recursive cascaded networks for unsupervised medical image registration [C]//*2019 IEEE/CVF International Conference on Computer Vision*. Seoul: IEEE, 2019: 10599-10609.

[18] DALCA A V, BALAKRISHNAN G, GUTTAG J, et al. Unsupervised learning for fast probabilistic diffeomorphic registration [M]//Medical image computing and computer assisted intervention — MICCAI 2018. Cham: Springer: 729-738.

[19] YANG X, KWITT R, STYNER M, et al. Quicksilver: Fast predictive image registration — A deep learning approach [J]. *NeuroImage*, 2017, **158**: 378-396.

[20] KREBS J, DELINGETTE H, MAILHE B, et al. Learning a probabilistic model for diffeomorphic registration [J]. *IEEE Transactions on Medical Imaging*, 2019, **38**(9): 2165-2176.

[21] MOK T C W, CHUNG A C S. Fast symmetric diffeomorphic image registration with convolutional neural networks [C]//*2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition*. Seattle: IEEE, 2020: 4643-4652.

[22] MOK T C W, CHUNG A C S. Large deformation diffeomorphic image registration with Laplacian pyramid networks [M]//Medical image computing and computer assisted intervention – MICCAI 2020. Cham: Springer, 2020: 211-221.

[23] HU R X, WANG H K, RISTANIEMI T, et al. Lung CT image registration through landmark-constrained learning with convolutional neural network [C]//*2020 42nd Annual International Conference of the IEEE Engineering in Medicine & Biology Society*. Montreal: IEEE, 2020: 1368-1371.

[24] PUNN N S, AGARWAL S. Automated diagnosis of COVID-19 with limited posteroanterior chest X-ray images using fine-tuned deep neural networks [J]. *Applied Intelligence*, 2021, **51**: 2689-2702.

[25] AMIRI M, BROOKS R, RIVAZ H. Fine-tuning U-net for ultrasound image segmentation: Different layers, different outcomes [J]. *IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control*, 2020, **67**(12): 2510-2518.

[26] CHEN C, DOU Q, CHEN H, et al. Unsupervised bidirectional cross-modality adaptation via deeply synergistic image and feature alignment for medical image segmentation [J]. *IEEE Transactions on Medical Imaging*, 2020, **39**(7): 2494-2505.

[27] FERRANTE E, OKTAY O, GLOCKER B, et al. On the adaptability of unsupervised CNN-based deformable image registration to unseen image domains [M]//Machine learning in medical imaging. Cham: Springer, 2018: 294-302.

[28] XUE P, DONG E Q, JI H Z. Lung 4D CT image registration based on high-order Markov random field [J]. *IEEE Transactions on Medical Imaging*, 2020, **39**(4): 910-921.

[29] BALIK S, WEISS E, JAN N, et al. Evaluation of 4-dimensional computed tomography to 4-dimensional cone-beam computed tomography deformable image registration for lung cancer adaptive radiation therapy [J]. *International Journal of Radiation Oncology Biology Physics*, 2013, **86**(2): 372-379.

[30] CLARK K, VENDT B, SMITH K, et al. The cancer imaging archive (TCIA): Maintaining and operating a public information repository [J]. *Journal of Digital Imaging*, 2013, **26**(6): 1045-1057.

[31] HUGO G D, WEISS E, SLEEMAN W C, et al. A longitudinal four-dimensional computed tomography and cone beam computed tomography dataset for image-guided radiation therapy research in lung cancer [J]. *Medical Physics*, 2017, **44**(2): 762-771.

[32] ROMAN N O, SHEPHERD W, MUKHOPADHYAY N, et al. Interfractional positional variability of fiducial markers and primary tumors in locally advanced non-small-cell lung cancer during audiovisual biofeedback radiotherapy [J]. *International Journal of Radiation Oncology Biology Physics*, 2012, **83**(5): 1566-1572.

[33] DICE L R. Measures of the amount of ecologic association between species [J]. *Ecology*, 1945, **26**(3): 297-302.