Lung CT Image Registration through Landmark-constrained Learning with Convolutional Neural Network

Ruxue Hu, Hongkai Wang*, Tapani Ristaniemi, *Senior Member, IEEE*, Wentao Zhu*, Xiaobang Sun

***Abstract*—Accurate registration of lung computed tomography (CT) image is a significant task in thorax image analysis. Recently deep learning-based medical image registration methods develop fast and achieve promising performance on accuracy and speed. However, most of them learned the deformation field through intensity similarity but ignored the importance of aligning anatomical landmarks (e.g., the branch points of airway and vessels). Accurate alignment of anatomical landmarks is essential for obtaining anatomically correct registration. In this work, we propose landmark constrained learning with a convolutional neural network (CNN) for lung CT registration. Experimental results of 40 lung 3D CT images show that our method achieves 0.93 in terms of Dice index and 3.54 mm of landmark Euclidean distance on lung CT registration task, which outperforms state-of-the-art methods in registration accuracy.**

I. INTRODUCTION

Accurate registration of lung computed tomography (CT) image is the prerequisite for subsequent medical image analysis applications such as lung disease progression monitoring, multi-modal image fusion, 4D sequences motion correction for ventilation and perfusion estimation and so on [1]. Various researches have focused on this field over the past years. Medical image registration is to establish optimal point-to-point correspondences between a pair of medical images, and is commonly solved by optimization on a similarity metric. As robust and discriminative similarity metric always leads to more accurate correspondences in registration tasks, design of an optimal similarity metric is one of the most important part for achieving accurate registration.

In classical lung CT registration methods, similarity metric is defined as intensity-based, feature-based or the combining of the two similarity metrics, which the intensity similarity is usually measured as mutual information (MI), sum of squared differences (SSD) or local cross correlation (LCC), and anatomical features are commonly using landmark, contour, or surface [2-5]. Among feature-based similarity metrics, anatomical landmark is one of the most frequently used features, especially for fine anatomical structures registration like lobe fissures, lung vessels and airways. Rühaak *et al.* [6] achieved robust and accurate correspondence for pulmonary CT images through integrating the landmark information as a least-squares penalty into registration optimization. Schmidt-Richberg *et al.* [7] proposed landmark-driven parameter optimization for thoracic 4D CT images registration. Polzin *et al.* [3] developed an automatically landmark detection scheme and incorporated the detected landmark information into Thin-Plate-Spline method for accurate lung CT registration. However, most of these landmark incorporated registration methods are based on conventional iterative optimization registration scheme, which is always time consuming and computationally intensive.

Recently, deep learning shows good potential for different medical image registration tasks including lung CT registration. Krebs *et al.* [8] proposed a probabilistic auto encoder framework for cardiac registration and trained the model with LCC intensity similarity metric. Dalca *et al.* [9] developed a convolutional neural network (CNN) based probabilistic diffeomorphic registration for brain task and optimized the registration model with SSD intensity similarity metric. Fan *et al.* [10] constructed a dual-supervised CNN for 3D brain registration task and optimized the network with multi-scale image intensity similarity metric. These deep learning registration methods always achieved sub-second registration, which outperform conventional registration methods on efficiency. Even though, so far, most of them optimized based on intensity similarity metric and rarely consider the importance of anatomical landmark alignment.

Therefore, we propose a landmark-constrained CNN model for lung CT registration. Inspired by Dalca *et al.* [9], we substantially extend a CNN based probabilistic diffeomorphic registration method with extra landmark constraint. The probabilistic diffeomorphic registration method enables large deformation and meanwhile the deformation is differentiable, invertible and topology preserving. Through adding the landmark constraint for learning, we achieve better registration accuracy of fine anatomical structures like the branch points of airways. Our method only requires expert-defined landmarks to constrain the network optimization at the training stage, and it does not need any further landmark input when we apply the trained model for image registration task in the testing stage. Therefore, the method is semi-supervised for training while fully automated for testing.

II. METHOD

We construct a CNN and take it as a probability registration model for achieving an optimal deformation field. Through stochastic gradient decent optimization of the CNN, the parameters of the probability registration model are estimated. The network architecture is based on the VoxelMorph model [9] and we add additional landmark

Ruxue Hu, Hongkai Wang and Xiaobang Sun are with the School of Biomedical Engineering, Dalian University of Technology, Dalian, China.

Ruxue Hu, Tapani Ristaniemi and Xiaobang Sun are with Faculty of Information Technology, University of Jyväskylä, Jyväskylä, Finland.

Wentao Zhu is with the Zhejiang Lab, Hangzhou, Zhejiang, China.

*Corresponding authors: Hongkai Wang (wang.hongkai@dlut.edu.cn) and Wentao Zhu (wentao.zhu@zhejianglab.com).

978-1-7281-1990-8/20/$31.00 ©2020 IEEE

constraint to ensure accurate alignment of crucial anatomical key points.

A. Probability Registration Model

3D image registration is to find an optimal deformation field ψ, which can warp a moving image $m \in R^3$ to a fixed image $f \in R^3$. As mentioned in [9], in order to ensure the deformation field is differentiable, invertible and topology preserving, we apply a stationary velocity field (SVF) to compute the deformation field ψ [11]. Let v be the SVF and then the deformation field ψ is computed according to ordinary differential equation [12],

$$\frac{\partial \psi^{(t)}}{\partial t} = v\left(\psi^{(t)}\right), \tag{1}$$

which is realized with scaling and squaring technique [13].

The prior probability of the SVF v is modeled as

$$p(v) = N(v; 0, \Sigma_v), \tag{2}$$

where $N(\cdot; \mu, \Sigma)$ is the multivariate normal distribution with mean μ and covariance Σ. From the SVF v, we derive the deformation field ψ_v. Then the registration of image m and landmark l_f can be represented as $m \circ \psi_v$ and $l_f \circ \psi_v$. According to the above, we construct the probability model of the images and landmarks registration as

$$p(f|v; m) = N\left(f; m \circ \psi_v, \sigma^2 \mathrm{II}\right), \tag{3}$$

$$p\left(l_m|v; l_f\right) = N\left(l_m; l_f \circ \psi_v, \sigma_l^2 \mathrm{II}\right), \tag{4}$$

where l_m is the landmark set of the moving image m, l_f is the landmark set of the fixed image f, σ^2 captures the variance of additive image noise, σ_l^2 captures the variance of additive displacement noise. We assume the true posterior probability of the model as $p\left(v|f, l_m; m, l_f\right)$, and the parameterized posterior probability is modeled as

$$q_\theta(v|f; m) = N\left(v; \mu_{v|m,f}, \Sigma_{v|m,f}\right), \tag{5}$$

where θ is the learnable parameters of our network and $\mu_{v|m,f}$ and $\Sigma_{v|m,f}$ of v is the output of our network.

For solving the probability registration model, we aims to approximate the true posterior probability with the parameterized posterior probability. According to the variance inference approach [14], we estimate the parameters of the model through minimization of the Kullback-Leibler divergence between the true and approximate posterior deformation probability as

$$\begin{aligned}
&\min_\theta KL\left[q_\theta(v|f; m) \parallel p\left(v|f, l_m; m, l_f\right)\right] \\
&= \min_\theta KL[q_\theta(v|f; m) \parallel p(v)] \\
&\quad - E_q[\log p(f|v; m)] - E_q\left[\log p\left(l_m|v; l_f\right)\right].
\end{aligned} \tag{6}$$

Similar to [9], we define the Laplacian of a neighborhood graph on the voxel grid as $L = D - N$, and D is the graph degree matrix, N is a voxel neighborhood adjacency matrix. We define a precision matrix $\Lambda_v = \Sigma_v^{-1} = \lambda L$ to ensure spatial smoothness, which λ sets the scale of SVF v. Then from the minimization of Kullback-Leibler divergence, we can arrive the loss function of our model as

$$\begin{aligned}
&\mathcal{L}\left(\theta; f, l_m, m, l_f\right) \\
&= \frac{1}{2}\left[tr\left(\lambda D \Sigma_{v|m;f} - \log \Sigma_{v|m;f}\right) + \mu_{v|m,f}^T \Lambda_v \mu_{v|m,f}\right] \\
&\quad + \frac{1}{2\sigma^2}\|f - m \circ \psi_v\|^2 + \frac{1}{2\sigma_l^2}\left\|l_m - l_f \circ \psi_v\right\|^2,
\end{aligned} \tag{7}$$

where the first term constrain the posterior close to the prior $p(v)$, the second term constrain the warped image $m \circ \psi_v$ similar to the fixed image f, the third term constrain the warped landmarks $l_f \circ \psi_v$ similar to the landmarks l_m, λ, σ^2 and σ_l^2 are the hyper-parameters of our network.

B. Network Architecture

As shown in Figure. 1, our network consists of a U-net similar structure, a sample layer, an integration layer and a spatial transform layer. It contains four down-sample convolutional layers, three up-sample convolutional layers, three convolutional layers and three skip connections with copy operation in the U-net similar structure. All activations of the network are LeakyReLU and sizes of the convolution kernels are $3 \times 3 \times 3$. The U-net inputs the fixed image f and the moving image m, and outputs $\mu_{v|m,f}$ and $\Sigma_{v|m;f}$ of the velocity field v. The sample layer is implemented following the re-parameterization trick in [9], and is to sample out the velocity field v from the U-net's output as $v \sim N\left(\mu_{v|m,f}, \Sigma_{v|m,f}\right)$. The integration layer is realized with scaling and squaring technique [13], and is to derive the deformation field ψ_v from the velocity field v. The spatial transform layer is constructed through the spatial transformer networks [14], which is to register the landmark of fixed image l_f as $l_f \circ \psi_v$ and the moving image m as $m \circ \psi_v$ with the deformation field ψ_v.

Actually, through the deformation field ψ_v we derive the displacement vector, which specifies the vector offset from the fixed image f to the moving image m for each voxel. Therefore, in the spatial transform layer, for each voxel location p in the fixed image f, we can estimate the correspondence voxel location q with the location shift $\psi(p)$ as

$$q = p + \psi(p). \tag{8}$$

As the definition of the voxel value is on integer location, we linearly interpolate at the neighboring voxels of q to get the correspondence voxel value $m \circ \psi_v$ as

$$m \circ \psi_v = \sum_{a \epsilon \mathcal{Z}(q)} m(a) \prod_{d \in \{x,y,z\}} (1 - |q_d - a_d|), \tag{9}$$

where $\mathcal{Z}(q)$ are the voxel neighbors of q. For the landmark l_f in the fixed image f, we estimate the corresponding landmark location $l_f \circ \psi_v$ with the location shift $\psi\left(l_f\right)$ as

$$l_f \circ \psi_v = l_f + \psi\left(l_f\right). \tag{10}$$

III. EXPERIMENT

A. Data

We collect 40 healthy adults' CT images, and 11 expert-annotated landmarks of the lung region are included with per image. The landmarks include bifurcation points of trachea, left and right superior lobar bronchus, right middle lobar

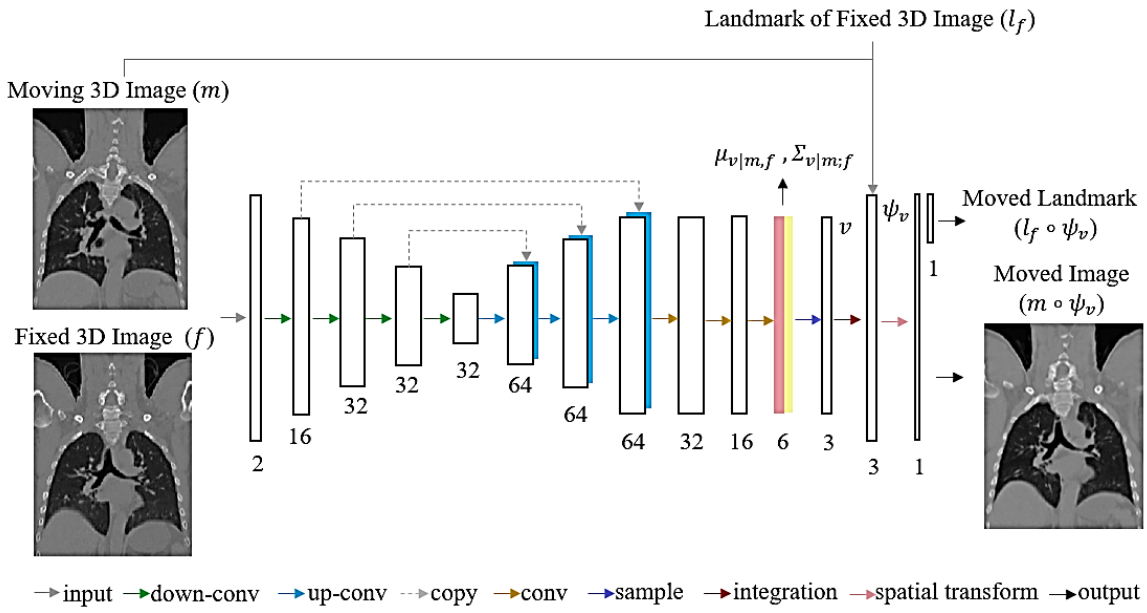

Figure 1. Network architecture

TABLE I
AVERAGE TARGET REGISTRATION ERRORS (MM) ON LANDMARKS (MEAN ± STANDARD DEVIATION) FOR EACH OF THE IMAGE PAIRS IN THE TEST DATASET

Image Pair	Elastix (BSpline)	ANTs (SyN)	VoxelMorph	Our Method
1	3.88 ± 2.18	4.11 ± 2.56	4.28 ± 3.15	**3.56 ± 2.32**
2	3.24 ± 2.03	4.05 ± 1.65	4.75 ± 1.69	**3.24 ± 1.87**
3	3.4 ± 1.95	3.41 ± 1.67	3.54 ± 1.81	**3.13 ± 1.16**
4	3.78 ± 1.06	3.98 ± 1.11	3.44 ± 1.24	**3.20 ± 0.84**
5	7.82 ± 2.27	7.62 ± 2.73	9.14 ± 4.49	**4.60 ± 2.10**
6	3.95 ± 1.79	4.05 ± 2.07	5.59 ± 3.07	**3.15 ± 1.77**
7	5.52 ± 3.26	7.31 ± 2.61	6.86 ± 3.69	**4.18 ± 0.91**
8	4.27 ± 1.62	4.15 ± 1.71	4.59 ± 2.73	**3.39 ± 1.97**
9	5.31 ± 1.32	5.77 ± 2.26	5.14 ± 1.87	**4.55 ± 1.22**
10	3.49 ± 2.33	3.70 ± 2.32	4.84 ± 2.83	**2.38 ± 1.13**
All	4.47 ± 2.44	4.82 ± 2.56	5.22 ± 3.15	**3.54 ± 1.68**

bronchus, pulmonary artery, left and right pulmonary artery, and four points on trachea with equal distance intervals. The images are resampled to 2mm isotropic voxel, affine aligned using Elastix software package and cropped to $160 \times 112 \times 208$. We randomly select 30 for training and use the remaining 10 for testing.

B. Configuration

All experiments are implemented on a computer with Intel(R) Xeon(R) CPU E5-2687W v4 @ 3.00GHz, GPU NVIDIA Tesla P40. We use Keras deep learning framework in the experiment. The network is optimized with Adam. The network parameters is initialized with an initialization model from [9]. The hyper-parameters are set as the initial learning rate $\eta = 0.0001$, $\lambda = 10$, $\sigma^2 = 0.02$ and $\sigma_i^2 = 1$.

C. Evaluation

We evaluate the registration performance on both accuracy and speed. We evaluate the accuracy through average target registration error (TRE) on landmarks and average Dice [15] on images. We propagate the landmark with the deformation field through the spatial transform layer and measure the TRE on landmarks using Euclidean distance. Similarly, we propagate the segmentation map of lungs with the deformation field through the spatial transform layer and compute the volume overlap with Dice. The TRE and Dice are defined as

$$TRE = \sqrt{(f_x - r_x)^2 + (f_y - r_y)^2 + (f_z - r_z)^2}, \quad (11)$$

$$Dice = 2\frac{|R_f \cap R_r|}{|R_f| + |R_r|}, \quad (12)$$

where (f_x, f_y, f_z) represents landmark location coordinate of the subject, (r_x, r_y, r_z) represents registered landmark location coordinate, R_f represents expert-segmented lung regions, R_r represents registered segmented region, $|\cdot|$ represents region volume, $\cap$ represents region overlapping. We evaluate the registration speed with the average registration time.

The compared methods include the Elastix software package with mutual information metric and BSpline spatial transform [16], the ANTs software package with Symmetric Normalization (SyN) [17] and VoxelMorph [9] which is the baseline method that we extend. As shown in Table Ⅰ, we calculate the average TRE on landmarks for 10 subjects in the test dataset for each method. It shows that our method significantly improves registration accuracy on average TRE as comparison to other three methods. Meanwhile, our method also shows robustness and generalization among subjects in test dataset. As in Table Ⅱ, we measure the average Dice on 10 subjects in the test dataset and the average registration time for each method. It demonstrates that our method outperforms the other three methods on accuracy with average Dice metric as well. Noticeably, our method obtains slightly better Dice index but much better landmark distance than the baseline method, which indicates our method successfully aligns the crucial anatomical landmarks. Moreover, our method takes less than a second with GPU, which is much faster than the conventional CPU-based registration methods (Elastix and ANTs).

TABLE II
AVERAGE DICE SCORES OVER LUNG REGION AND AVERAGE REGISTRATION TIME

Method	Avg. Dice	GPU sec	CPU sec
Elastix (BSpline)	0.91	–	49
ANTs (SyN)	0.90	–	443
VoxelMorph	0.92	**0.5**	–
Our Method	**0.93**	**0.5**	–

IV. CONCLUSIONS

In this paper, we construct a landmark-constrained learning architecture and apply it on lung CT registration task. With the proposed approach, we achieve a smooth, large deformation for accurate lung CT registration. Especially, it assures more accurate alignment of key landmarks which are anatomically meaningful. As demonstrated in our experiments, the proposed approach improves the registration accuracy comparing with the baseline method and two classical methods. Besides, it does not require landmark input for image registration tasks on the trained model and the registration speed is sub-second fast. It is also meaningful to note that this landmark-constrained learning is effective for fine geometric structure registration and can be expanded to registration applications other than lung CT images. For further research, we plan to collect a larger lung CT image dataset and apply multi-resolution scheme on the proposed architecture to further improve the accuracy performance.

ACKNOWLEDGMENT

This study is supported by the general program of the National Natural Science Fund of China (No. 81971693, 81401475), and the Fundamental Research Funds for the Central Universities (DUT19JC01).

REFERENCES

[1] M. P. Heinrich, "Deformable lung registration for pulmonary image analysis of MRI and CT scans," Ph.D. dissertation, Oxford University, UK, 2013.

[2] J. Ehrhardt, R. Werner, A. Schmidt-Richberg, and H. Handels, "Automatic landmark detection and non-linear landmark-and surface-based registration of lung CT images," *Medical Image Analysis for the Clinic-A Grand Challenge, MICCAI,* vol. 2010, pp. 165-174, 2010.

[3] T. Polzin, J. Rühaak, R. Werner, J. Strehlow, S. Heldmann, H. Handels, *et al.*, "Combining automatic landmark detection and variational methods for lung CT registration," in *Fifth International Workshop on Pulmonary Image Analysis*, 2013, pp. 85-96.

[4] B. Li, G. E. Christensen, E. A. Hoffman, G. McLennan, and J. M. Reinhardt, "Establishing a normative atlas of the human lung: intersubject warping and registration of volumetric CT images," *Academic radiology,* vol. 10, pp. 255-265, 2003.

[5] D. L. Hill, P. G. Batchelor, M. Holden, and D. J. Hawkes, "Medical image registration," *Physics in medicine & biology,* vol. 46, p. R1, 2001.

[6] J. Rühaak, T. Polzin, S. Heldmann, I. J. A. Simpson, and M. P. Heinrich, "Estimation of Large Motion in Lung CT by Integrating Regularized Keypoint Correspondences into Dense Deformable Registration," *IEEE Transactions on Medical Imaging,* vol. 36, pp. 1746-1757, 2017.

[7] A. Schmidt-Richberg, R. Werner, J. Ehrhardt, J. C. Wolf, and H. Handels, "Landmark-driven Parameter Optimization for non-linear Image Registration," in *Medical Imaging 2011: Image Processing*, 2011, p. 79620T.

[8] J. Krebs, H. Delingette, B. Mailhé, N. Ayache, and T. Mansi, "Learning a probabilistic model for diffeomorphic registration," *IEEE transactions on medical imaging,* vol. 38, pp. 2165-2176, 2019.

[9] A. V. Dalca, G. Balakrishnan, J. Guttag, and M. R. Sabuncu, "Unsupervised learning of probabilistic diffeomorphic registration for images and surfaces," *Medical image analysis,* vol. 57, pp. 226-236, 2019.

[10] J. Fan, X. Cao, P.-T. Yap, and D. Shen, "BIRNet: Brain image registration using dual-supervised fully convolutional networks," *Medical image analysis,* vol. 54, pp. 193-206, 2019.

[11] J. Ashburner, "A fast diffeomorphic image registration algorithm," vol. 38, pp. 95-113, 2007.

[12] C. Moler and C. Van Loan, "Nineteen dubious ways to compute the exponential of a matrix, twenty-five years later," *SIAM review,* vol. 45, pp. 3-49, 2003.

[13] V. Arsigny, O. Commowick, X. Pennec, and N. Ayache, "A log-euclidean framework for statistics on diffeomorphisms," in *International Conference on Medical Image Computing and Computer-Assisted Intervention*, 2006, pp. 924-931.

[14] M. Jaderberg, K. Simonyan, and A. Zisserman, "Spatial transformer networks," in *Advances in neural information processing systems*, 2015, pp. 2017-2025.

[15] L. R. Dice, "Measures of the Amount of Ecologic Association Between Species," *Ecology,* vol. 26, pp. 297-302, 1945.

[16] S. K. Balci, P. Golland, M. Shenton, and W. M. Wells, "Free-Form B-spline Deformation Model for Groupwise Registration," in *Medical image computing and computer-assisted intervention: MICCAI... International Conference on Medical Image Computing and Computer-Assisted Intervention*, 2006, pp. 23-30.

[17] B. B. Avants, C. L. Epstein, M. Grossman, and J. C. Gee, "Symmetric diffeomorphic image registration with cross-correlation: evaluating automated labeling of elderly and neurodegenerative brain," *Medical image analysis,* vol. 12, pp. 26-41, 2008.