

Article

Intra-Patient Lung CT Registration through Large Deformation Decomposition and Attention-Guided Refinement

Jing Zou ¹, Jia Liu ², Kup-Sze Choi ¹ and Jing Qin ^{1,*}

¹ Center for Smart Health, School of Nursing, the Hong Kong Polytechnic University, Hong Kong, China; jing9.zou@connect.polyu.hk (J.Z.); jia.liu@siat.ac.cn (J.L.); thomask.s.choi@polyu.edu.hk (K.-S.C.)

² Shenzhen Institutes of Advanced Technology, Chinese Academy of Sciences, Shenzhen 518055, China

* Correspondence: harry.qin@polyu.edu.hk

Abstract: Deformable lung CT image registration is an essential task for computer-assisted interventions and other clinical applications, especially when organ motion is involved. While deep-learning-based image registration methods have recently achieved promising results by inferring deformation fields in an end-to-end manner, large and irregular deformations caused by organ motion still pose a significant challenge. In this paper, we present a method for registering lung CT images that is tailored to the specific patient being imaged. To address the challenge of large deformations between the source and target images, we break the deformation down into multiple continuous intermediate fields. These fields are then combined to create a spatio-temporal motion field. We further refine this field using a self-attention layer that aggregates information along motion trajectories. By leveraging temporal information from a respiratory cycle, our proposed methods can generate intermediate images that facilitate image-guided tumor tracking. We evaluated our approach extensively on a public dataset, and our numerical and visual results demonstrate the effectiveness of the proposed method.

Keywords: deformable medical image registration; intra-patient; lung CT; deformation field decomposition; attention layer refinement



Citation: Zou, J.; Liu, J.; Choi, K.-S.;

Qin, J. Intra-Patient Lung CT

Registration through Large

Deformation Decomposition and

Attention-Guided Refinement.

Bioengineering **2023**, *10*, 562.

[https://doi.org/10.3390/](https://doi.org/10.3390/bioengineering10050562)

[bioengineering10050562](https://doi.org/10.3390/bioengineering10050562)

Academic Editors: Ming Liu,

Liquan Dong, Qingliang Jiao

and Antonio Brunetti

Received: 21 March 2023

Revised: 28 April 2023

Accepted: 6 May 2023

Published: 8 May 2023



Copyright: © 2023 by the authors.

Licensee MDPI, Basel, Switzerland.

This article is an open access article

distributed under the terms and

conditions of the Creative Commons

Attribution (CC BY) license ([https://](https://creativecommons.org/licenses/by/4.0/)

[creativecommons.org/licenses/by/](https://creativecommons.org/licenses/by/4.0/)

[4.0/](https://creativecommons.org/licenses/by/4.0/)).

1. Introduction

Deformable medical image registration (DMIR) is a critical task in various clinical applications, including surgery planning [1], diagnosis [2], radiation therapy [3,4], and longitudinal studies [5]. It involves aligning multiple images acquired at different times, from different patients, or from different positions. Among various medical imaging registration challenges, lung CT registration is particularly difficult due to the large and non-rigid deformations caused by organ motion, respiratory cycles, and pathological changes [6]. The accurate and efficient registration of lung CT images is crucial for disease diagnosis, monitoring, and radiation therapy planning. Therefore, developing a patient-specific lung CT image registration method that addresses these challenges is of paramount importance.

The registration process involves finding a transformation that aligns the anatomical structures in a source image with those in a target image. In early studies, traditional registration methods have been proposed, such as Large Diffeomorphic Distance Metric Mapping (LDDMM) [7–9] and symmetric methods [10–12]. However, these traditional methods are often computationally expensive and time-consuming as they formulate the registration problem as an independent pair-wise iterative optimization problem. The registration process becomes significantly time-consuming when dealing with image pairs containing significant deformations in anatomical structures. This limitation prevents such techniques from being used in many clinical scenarios where real-time registration is required. Moreover, these methods often struggle to handle large and complex deformations between images, such as those caused by organ motion or respiration.

Deep learning techniques have been utilized to solve deformable registration problems in recent years. Supervised methods were the first to be introduced for performing image registration, where a ground truth deformation field is necessary for training. However, these ground truth deformation fields are difficult to obtain, especially for medical image registration tasks. Then, researchers turned to unsupervised methods and semi-supervised methods, which eliminated the requirement of ground truth. Many unsupervised methods and semi-supervised methods have achieved remarkable success [13–20].

Despite the progress made in the use of deep learning techniques for deformable image registration, accurately estimating large and irregular deformations caused by respiration remains a challenging task. This is particularly important in the context of medical image registration, as tumors and sensitive structures in the thorax can move more than 20 mm [21] during respiration, making it difficult to achieve accurate registration. Figure 1 shows a large deformation in the diaphragm and near the heart area.

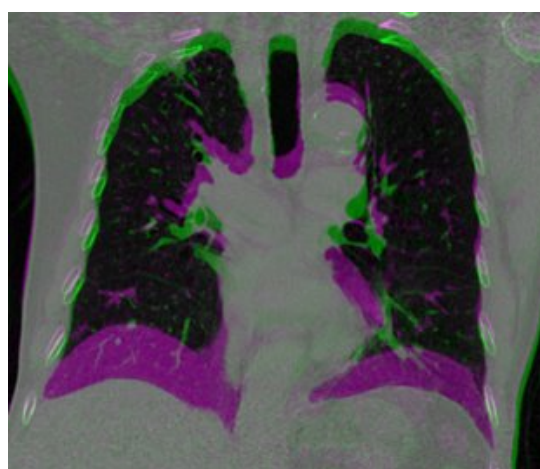


Figure 1. A composite image by overlaying the inspiration image onto the expiration image. The large differences are presented as magenta color.

We propose a patient-specific method for deformable lung CT image registration by decomposing the large deformation fields into several continuous intermediate fields. This approach allows us to better capture the complex and large deformations that occur in the lung and improves the accuracy of the registration results. Our proposed approach takes advantage of the temporal characteristics within a respiratory cycle to create intermediate images, which can be valuable in applications such as tumor tracking within image-guided systems. During the training process, both the extreme phase images and the intermediate images are utilized as training data. After the network has completed the training process, it is capable of estimating a deformation field directly without relying on any intermediary images.

We conduct an evaluation of our proposed method on a publicly available dataset, and the results of our experiments show that our approach is effective in enhancing the accuracy and efficiency of lung CT registration. The proposed method has potential applications in radiation therapy planning, disease monitoring, and other clinical scenarios that require accurate and efficient lung CT registration.

2. Related Works

2.1. Pair-Wise Optimization

Traditional image registration methods have been widely studied in medical image registration. These methods usually treat registration as an independent pair-wise iterative optimization problem, where the objective is to find the optimal transformation parameters that map one image to another. Many toolkits are open to the public for medical image analysis, such as Advanced Normalization Tools (ANTs) [22], Simple Elastix [23], and NiftyReg [24]. One of the most well-known traditional methods is Large Diffeomorphic Distance Metric Map-

ping (LDDMM) [7–9]. The registration process is represented as a partial differential equation that governs the deformation field between two images and describes the transformation that aligns one image with another. LDDMM is able to handle large deformations and topological changes in the image, making it suitable for registration tasks in medical imaging. Another example is Symmetric Normalization (SyN) [10–12], which uses a symmetric diffeomorphic transformation to optimize the similarity between two images.

However, these traditional methods have several limitations. One of the major drawbacks is that they are usually computationally expensive and time-consuming, and they are particularly problematic when dealing with image pairs that exhibit significant deformations in anatomical structures. This is because they need to solve a complex optimization problem for each pair of images independently. Moreover, traditional methods often require manual initialization, which is labor-intensive and time-consuming. Another issue is that traditional methods are sensitive to the choice of similarity measures and regularization terms, which require expert knowledge to tune properly. Recent studies have addressed the aforementioned limitations by focusing on the development of deep-learning-based methods for medical image registration. These methods aim to directly learn the deformation field from the input images, eliminating the need for explicit optimization.

2.2. Learning-Based Registration

In recent years, image registration methods based on machine learning have gained popularity. These approaches can be broadly classified into three categories based on the degree of supervision required: supervised, unsupervised, and weakly supervised methods.

Supervised methods for image registration [25–27] rely on ground truth deformation fields, which can be either real or synthetic. The real deformation fields are obtained through traditional registration techniques, while synthetic deformation fields are generated using statistical models or random transformations. These fields are used as a reference to train the registration algorithm. The quality and quantity of the ground truth fields significantly affect the performance of the model.

Unsupervised methods for image registration do not require annotated training data and instead learn the deformation fields by minimizing the dissimilarity between the target and transformed source images. Balakrishnan et al. [20] proposed an unsupervised method and adopted a Spatial Transformer Network (STN) [28] for image transformation. De Vos et al. [14] presented an unsupervised multi-stage network for affine and deformable registration adopting a coarse-to-fine strategy, while CycleMorph with cycle-consistency [17], inverse-consistency [29,30], and symmetric pyramid network [31] are proposed to improve the registration performance. However, these methods are trained in a bi-directional manner, which highly increases the computational complexity and training time.

The semi-supervised networks [32,33] leverage the mask or landmark information as supervisions during the training. In a popular network, VoxelMorph [13], a semi-supervised strategy, was utilized by supervising the learning of the deformation field with brain masks. For lung CT image registration, the network in [16] leverages lung masks and landmarks for semi-supervised learning. However, in the medical image domain, acquiring masks and labels is challenging and costly.

2.3. Self-Attention

Self-attention [34] is a mechanism in deep learning that enables a network to selectively focus on different parts of an input sequence or image, allowing it to encode relevant information and discard irrelevant information. Several studies have explored the use of self-attention in medical image registration, which has shown promising results.

Chen et al. [35] introduced a novel architecture for medical image registration, named ViT-V-Net, which combines ConvNet and Transformer models for brain MRI registration. They incorporated the Vision Transformer (ViT) [36] to learn high-level features, and found that replacing the VoxelMorph backbone with ViT-V-Net improved the registration

accuracy. They further extended ViT-V-Net to TransMorph [37], a hybrid Transformer-ConvNet framework that uses the Swin Transformer [38] as the encoder to learn spatial transformations between input images. After processing the data from the Transformer encoder, a decoder built with ConvNet generated the dense deformation field. In another work, Liu et al. [39] proposed multi-organ PET and CT multi-modality registration based on a pre-trained 2D Transformer model.

One advantage of using self-attention in medical image registration is its ability to capture global dependencies and long-range interactions, which are particularly relevant in medical imaging where structures can undergo large deformations. Additionally, self-attention can capture complex spatial relationships between different regions of the image, improving the accuracy of the registration. However, one disadvantage of self-attention-based registration methods is their relatively high computational cost. Self-attention requires the computation of a pairwise similarity matrix between all feature vectors in the input sequence, which can be computationally expensive for large input sizes. In this work, we alleviate this disadvantage by utilizing only one self-attention layer for spatial-temporal field refinement, rather than using the whole Transformer architecture [34].

3. Methodology

3.1. Problem Statement

Our objective is to obtain the deformation field $\Phi \in \mathbb{R}^{H \times W \times C \times 3}$ that represents the spatial transformation between the source lung CT image I_S and the target lung CT image I_T , where I_S and I_T are $H \times W \times C$ tensors. To achieve this, we utilize a deep neural network f , which takes I_S and I_T as inputs and generates Φ . The goal is to solve the following problem:

$$\operatorname{argmin}_{f \in \mathcal{F}} \ell(I_T, I_S \circ \Phi_f) + \lambda \mathcal{R}(\Phi_f), \quad (1)$$

where \mathcal{F} represents the function space of f and Φ_f represents the deformation field obtained using f for the given inputs. The term $I_S \circ \Phi_f$ represents the warped image obtained by applying Φ_f on I_S . The loss function ℓ is used to measure the difference between I_T and $I_S \circ \Phi_f$. Additionally, we introduce a regularization term $\mathcal{R}(\Phi_f)$ with a hyperparameter λ to control its influence on the optimization.

While the current training approach is effective for small deformations in lung CT images [7], it fails to produce accurate results when dealing with large and irregular deformations where pixels undergo significant transformations. To address this limitation, our method decomposes the deformation field into several small ones and gradually improves them using an attention layer. Figure 2 provides an overview of our proposed approach.

3.2. Decomposition

To address the issue of large and irregular deformations in lung CT image registration, we propose a method that decomposes the deformation field Φ into multiple intermediate fields u_i , which describe the direction and distance of voxel movement from I_S to I_T . We assume that the movement of each voxel is along a smooth line, and therefore, we use linear interpolation to obtain the intermediate fields u_i by incrementally moving the voxels along the deformation path.

$$u_i = \Phi/n, \quad (2)$$

n represents the number of phases in a respiratory cycle.

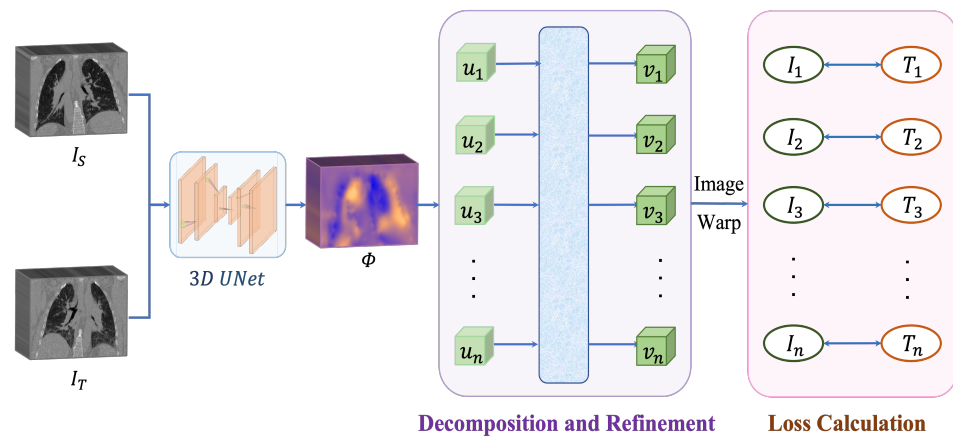


Figure 2. The illustrative pipeline of our method. The 3D UNet takes the source image and the target image as inputs and outputs a deformation field. Then, the deformation field is decomposed into several intermediate deformation fields, these deformation fields are refined through the attention layer. Finally, the network is updated by minimizing the similarity losses.

3.3. Refinement

The linear interpolation used for decomposing the deformation field assumes a homogeneous displacement of each voxel, which may not always hold true in practical scenarios. To address this, we refine the small deformation fields by concatenating them to create a spatio-temporal motion field U that contains both spatial and temporal information of the respiratory motion. We then pass this motion field U through a self-attention layer to obtain a refined field V , which is decomposed again to generate refined intermediate fields v_i . Using these refined fields, we warp the source image I_S to generate intermediate images I_n , which are compared with ground truth intermediate images T_n to calculate the loss.

Attention Layer

Self-attention is a mechanism commonly used in deep learning models that enables the model to selectively weigh the importance of different parts of the input sequence in making predictions. It achieves this by calculating the dot product of a query matrix with a key matrix, applying a softmax activation function to the resulting scores, and then multiplying the output by a value matrix to obtain the final representation. This allows the model to capture global information from the input sequence in a more efficient and effective manner. The self-attention formula is given as follows:

$$\text{Attention}(Q, K, V) = \text{softmax}\left(\frac{QK^T}{\sqrt{d_k}}\right)V, \tag{3}$$

where Q , K , and V are matrices representing query, key, and value, respectively, with dimensions $n \times d_k$, where n is the sequence length and d_k is the dimensionality of the key and value vectors. The dot product QK^T is divided by $\sqrt{d_k}$ to mitigate the effects of vanishing gradients. Finally, the softmax function is applied to the scores to obtain attention weights, which are then used to weight the value vectors V to obtain the final output.

In our experiment, we construct an attention layer consist of Layer Norm (LN) operations [40] and residual connections [41]:

$$y = y + \text{Norm}(\text{Attention}(x)), \tag{4}$$

3.4. Network Update

In our experiments, we used Mean Square Error (MSE) as the similarity matrix. This metric calculates the average of the squared differences between the target image and the transformed source image. The MSE is computed by taking the sum of the squared

residuals and dividing it by the number of voxels in the images. Mathematically, it can be expressed as follows:

$$\ell = \frac{1}{|\Omega|} \sum_{p \in \Omega} (I_T(p) - I_S \circ (\Phi)(p))^2 \tag{5}$$

Here, $I_T(p)$ and $I_S \circ (\Phi)(p)$ represent the intensity values of the target image and the deformed source image, respectively, at location p . $|\Omega|$ is the total number of voxels in the image domain Ω ,

To ensure the smoothness and physical plausibility of the deformation field Φ , we apply regularization during training. A diffusion regularization term is imposed on the spatial gradients of the deformation field to penalize large displacements. The regularization term can be expressed as follows:

$$\mathcal{R}(\Phi) = \sum_{p \in \Omega} \|\nabla(\Phi(p))\|^2, \tag{6}$$

where $\|\cdot\|^2$ denotes the $L2$ norm of the vector. We use finite differences to approximate spatial gradients, such that $\frac{\partial(\Phi)(p)}{\partial x} \approx (\Phi)((p_x + 1, p_y, p_z)) - (\Phi)((p_x, p_y, p_z))$ and similarly for $\frac{\partial u(p)}{\partial y}$ and $\frac{\partial(\Phi)(p)}{\partial z}$. This regularization term encourages a smoother deformation field Φ that is more physically plausible.

The ultimate goal of our decomposition approach is to train the deep neural network f to perform deformable registration. To achieve this, we aim to solve the following optimization problem:

$$\operatorname{argmin}_{f \in \mathcal{F}} \sum_{t=1}^n \ell_t(T_t, I_S \circ (t\Phi_f/n)) + \lambda \mathcal{R}(\Phi_f). \tag{7}$$

where ℓ_t represents the different loss function at the decomposed stage and T_t is the target intermediate images.

4. Experiments

4.1. Dataset

DirLab 4DCT dataset [42] contains 100 lung CT volumes collected from 10 patients, each having 10 CT volumes, denoted by T_0, \dots, T_9 , respectively, sampled at 10 different respiratory phases during a whole respiration period. All images have an isotropic in-plane resolution of 256×256 from 4DCT1 to 4DCT5 and of 512×512 from 4DCT6 to 4DCT10. The axial slices were acquired with a voxel size ranging from 0.97 mm to 1.98 mm, and the slice thickness and increment were 2.5 mm. This dataset also provides 300 anatomical landmarks for the volumes of T_0 and T_5 , representing the maximal inspiration and the maximal expiration, and 75 landmarks for the remaining volumes; these landmarks are useful for evaluating the registration performance. For additional information regarding the dataset used in our study, please refer to the provided link for the dataset (the dataset is released at the website <https://med.emory.edu/departments/radiation-oncology/research-laboratories/deformable-image-registration/downloads-and-reference-data/4dct.html>, accessed on 8 May 2023).

4.2. Data Pre-Processing

The data pre-process includes resampling, affine registration, and cropping. Firstly, all images are re-sampled into a voxel spacing of $1 \text{ mm} \times 1 \text{ mm} \times 2.5 \text{ mm}$. Then we performed affine registration for all cases by affine registering T_5 to T_0 , because in this study we focus on deformable registration. In the end, all the images are resized to have the same volume size of $240 \times 160 \times 96$.

4.3. Performance Metrics

We utilized two metrics to evaluate our registration algorithm: (1) Target Registration Error (TRE) and (2) the percentage of negative Jacobian Determinant values.

4.3.1. Target Registration Error (TRE)

The Target Registration Error (TRE) is used as the primary metric to assess the performance of our algorithm. TRE is a widely-used performance metric for landmark-based registration tasks, which measures the average Euclidean distance between landmarks in the target image and the corresponding landmarks in the registration output. A smaller TRE value indicates better registration performance.

$$TRE = \frac{1}{num} \sum_{i=1}^{num} \sqrt{((p - q) \cdot s)^2}, \quad (8)$$

where num is the total number of the landmarks used to evaluate the registration algorithms. $p \in \mathbb{R}^3$ is the landmark in the source image and $q \in \mathbb{R}^3$ is the landmark in the target image after transformation. s is the spacing of the images; in our experiments, the value is [1.0, 1.0, 2.5].

4.3.2. Jacobian Determinant

The Jacobian determinant is a measurement of the local spatial distortion of an image transformation. It quantifies how much the transformation expands or contracts the local volume of tissue. The calculation of the Jacobian determinant involves finding the partial derivatives of the transformation function with respect to each dimension of the image.

$$\det(J(i, j, k)) = \begin{vmatrix} \frac{\partial i}{\partial x} & \frac{\partial j}{\partial x} & \frac{\partial k}{\partial x} \\ \frac{\partial i}{\partial y} & \frac{\partial j}{\partial y} & \frac{\partial k}{\partial y} \\ \frac{\partial i}{\partial z} & \frac{\partial j}{\partial z} & \frac{\partial k}{\partial z} \end{vmatrix} \quad (9)$$

The determinant can be negative, positive, or zero. If the determinant of the Jacobian matrix is equal to 1, then there is no volume change during the transformation. If it is greater than 1, it means the volume has expanded, and if it is between 0 and 1, it means the volume has shrunk. However, a determinant less than or equal to 0 suggests that there is a singularity in the image after registration, which is a point where the image has been folded or distorted. By calculating the percentage of negative determinant values, we can quantify the quality of the deformation field.

4.4. Implementation Details

The algorithm is implemented with pytorch lightning framework on an RTX 3090 GPU. We employed 3D UNet [43] as the backbone network due to its high performance. We ran 600 epochs with the batch size of 1 which is set according to the memory of our GPU card. The hyperparameter before the regularization term in the loss function is set as 0.01. The optimizer we used is Adam [44] with an initial learning rate of 0.0001.

4.5. Experimental Results

4.5.1. Performance Improvement

We evaluate the performance of our proposed method by comparing it with five existing methods, namely BL [20], IL [14], VM [13], MAC [16], and CM [17], denoting the baseline and existing methods that use the iterative learning strategy, lung masks as the supervision, landmarks as the supervision, and the cycle consistency, respectively. To

ensure a fair comparison, we adopt the same backbone network (3D UNet) and learning settings for all methods.

The results of our experiments are presented in Table 1 obtained via cross-validation. Our proposed method achieved the best performance in terms of average target registration error (*Avg.* in the table) compared to the other five competitive methods. Specifically, our approach outperformed the second-best method (VM) by 8.0%, indicating its superior effectiveness. Moreover, our method yielded the best results for five out of the ten patients, and it also exhibited the lowest standard deviation (*Std.* in the table), indicating its reliability. These results demonstrate the effectiveness and robustness of our method.

Table 1. Numerical comparison of our proposed framework vs. other existing image registration algorithms. The numerical values reflect the TRE metric (mm) for the DIRLab 4DCT dataset. The best results are highlighted in bold font.

4DCT	BL	IL	VM	MAC	CM	Ours
01	2.29	4.24	2.28	2.35	2.01	2.20
02	1.97	2.19	2.19	2.19	1.97	1.75
03	2.62	2.52	2.43	2.68	2.98	2.18
04	2.70	3.16	2.59	2.66	2.48	2.56
05	4.16	4.74	3.90	4.12	4.03	4.45
06	2.52	3.18	3.14	3.32	2.75	2.64
07	4.18	4.24	3.76	3.61	4.34	3.39
08	6.02	5.93	5.79	6.27	6.77	4.24
09	3.43	2.95	3.07	3.37	2.96	2.88
10	5.40	5.36	4.69	4.75	5.35	4.80
<i>Avg.</i>	3.53	3.85	3.38	3.53	3.56	3.11
<i>Std.</i>	1.38	1.25	1.17	1.25	1.56	1.06

4.5.2. Statistical Analysis

We conducted a statistical significance test to confirm the improvement in performance using a non-parametric test—the Friedman test for multiple comparisons. The statistical analysis for the 4DCT dataset is reported in Figure 3. We can conclude that there is a statistically significant difference in performance compared to other methods with $\chi^2_{Friedman}(5) = 48.37$ and $p = 2.2 \times 10^{-7}$. In the figure, we also observed an outlier on the results using the MAC algorithm [16]. We do not find an outlier in our method, which demonstrates the effectiveness of our method again.

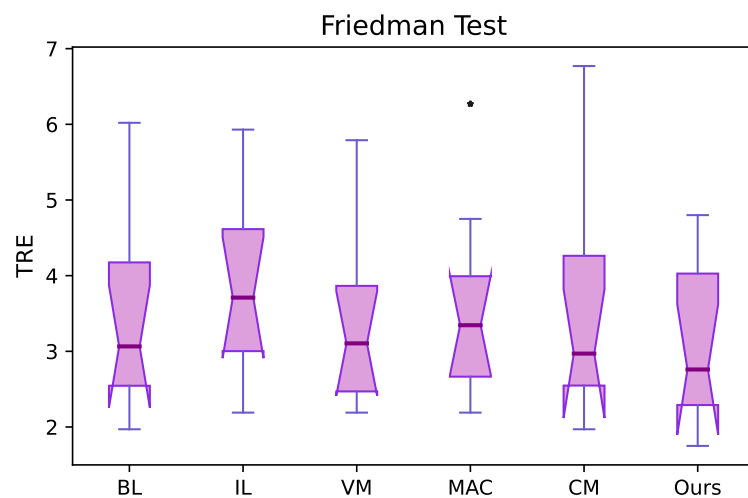


Figure 3. Friedman test. With $\chi^2_{Friedman}(5) = 48.37$ and $p = 2.2 \times 10^{-7}$, we can conclude that our method has significant improvement over other algorithms.

4.5.3. Evaluation on Jacobian Determinant

The percent of negative values of the Jacobian determinant is shown in Table 2. We can see that our method achieved the lowest percent (0.016) compared to other methods. Negative determinants indicate there is distortion or folding in the deformation field. Thus, the registration results of our method show less abrupt changes and outperform the others in terms of keeping the smoothness of the deformation field during registration.

Table 2. Comparison of our proposed method vs. other existing image registration algorithms on the percent of negative values of Jacobian determinant. The best results are highlighted in bold font.

Methods	BL	IL	VM	MAC	CM	Ours
$ J < 0$ (%)	0.023	0.018	0.020	0.020	0.022	0.016

4.5.4. Qualitative Evaluation

The visualization of results is shown in Figure 4. The first row displays the difference image before registration obtained by subtracting the target image from the source image. We can see that there are large differences in the diaphragm and near the heart area, which are illustrated as highlighted black areas. The second row shows the difference images after registration; we can observe that after registration, the large differences are all eliminated. The last row presents the deformation field from three different perspectives. In these images, the large deformations are presented in red and happen in the diaphragm near the heart area, which is in accordance with the large differences of the difference images in the first row.

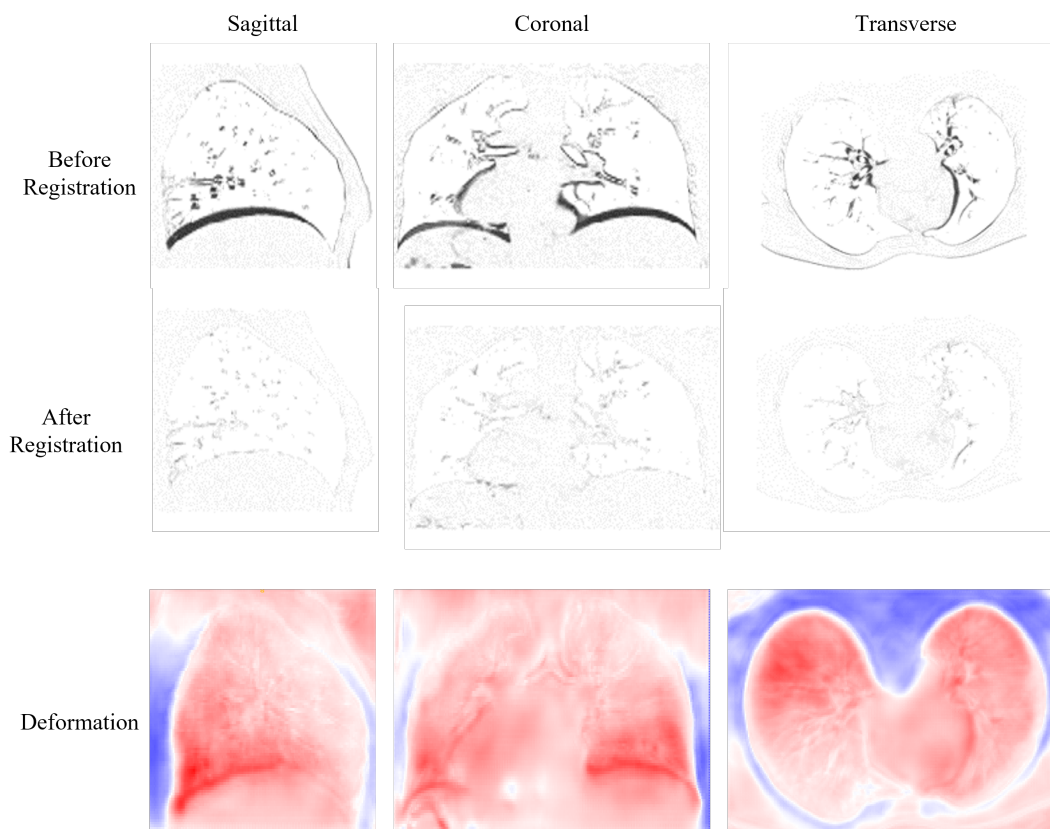


Figure 4. Visualization results for the difference images before and after registration, and deformation field from three perspectives.

Figure 5 shows the deformation field before and after refinement. The deformation field before refinement appears to be highly irregular, with visible artifacts and unrealistic

deformations. In contrast, the deformation field after refinement is significantly smoother and more natural-looking, with many of the aforementioned artifacts and deformations corrected.

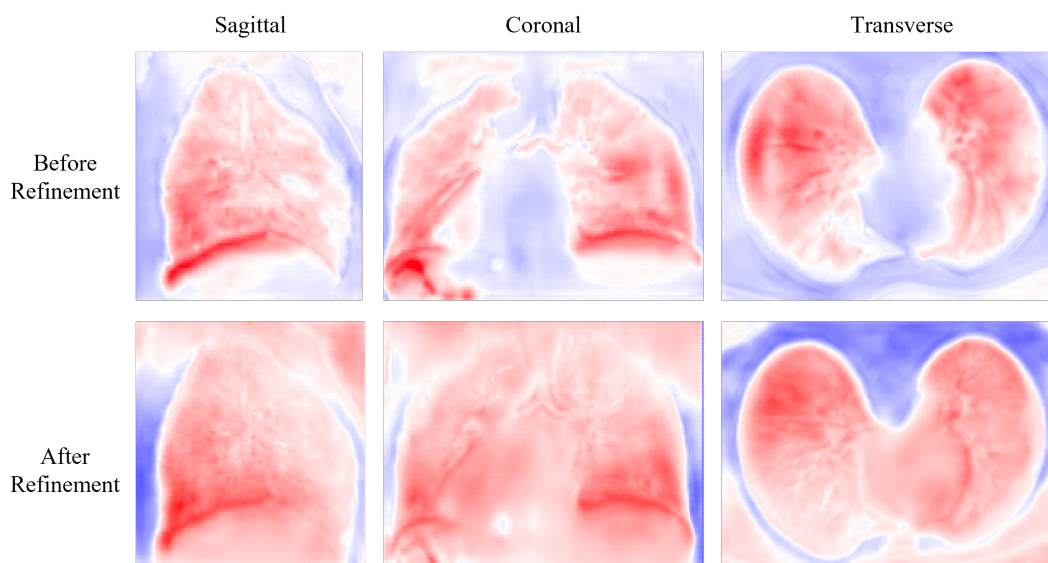


Figure 5. Visualization results for the slides of the deformation field before and after refinement.

5. Discussion

This paper proposes a method for deformable registration of lung CT scans in the same patient, which combines large deformation decomposition and attention-guided refinement. The results obtained using this approach show that it outperforms other state-of-the-art methods on a publicly available dataset. The method's ability to handle large deformations and variations in lung shape and texture is considered one of its key strengths. The use of attention mechanisms allows the method to focus on the most relevant features for registration and ignore noisy or irrelevant features, thereby increasing the algorithm's robustness and accuracy even in challenging cases.

Looking to the future, there are several avenues for further research that we intend to pursue. One important area for improvement is the addition of boundary constraints to our registration algorithm. We observed some boundary problems in the deformation field. In the future, we will solve this problem by adding a boundary constraint or applying a lung mask during registration. By incorporating boundary constraints, we can ensure that our registration method is more robust to variations in lung shape and better able to handle boundary artifacts that can arise in CT images.

In addition to boundary constraints, we also plan to explore other regularization techniques that may further improve the accuracy and robustness of our registration method. One promising approach is to incorporate information about lung function, such as pulmonary ventilation or perfusion, into our registration algorithm. This could enable us to better account for physiological variations between patients and to tailor the registration process to individual patient needs.

In addition to further improving our intra-patient lung CT registration algorithm, we also plan to evaluate its performance on other datasets and modalities. This will enable us to test the robustness of our approach under different conditions and to gain a better understanding of its potential applications in a variety of clinical settings. One of our future plans is to test the performance of our algorithm on datasets containing patients with diverse lung diseases such as lung cancer or chronic obstructive pulmonary disease (COPD). By testing our method on a broader range of patient populations, we can better assess its ability to handle variations in lung shape and texture that may be present in different disease states.

6. Conclusions

This paper explores a straightforward and efficient approach for learning the large deformation field in lung CT image registration, which is essential for image-guided navigation systems. The registration accuracy is improved by breaking down the large deformation field into smaller intermediate fields and then combining them through a composition process followed by refinement using an attention layer. The experimental results demonstrate that our method outperforms existing methods. Future work will involve investigating boundary constraints and regularizations, as well as conducting a large-scale evaluation of different imaging modalities.

Author Contributions: Conceptualization, J.Z. and J.Q.; methodology, J.Z. and J.L.; software, J.Z.; validation, J.Z.; formal analysis, J.Z.; resources, J.Q. and K.-S.C.; data curation, J.Z. and J.L.; writing—original draft preparation, J.Z.; writing—review and editing, J.L., J.Q. and K.-S.C.; visualization, J.Z.; supervision, K.-S.C. and J.Q.; project administration, J.Q.; funding acquisition, J.Q. All authors have read and agreed to the published version of the manuscript.

Funding: The work described in this paper is supported by grants of General Research Fund of Hong Kong Research Grants Council (Project Nos: 15218521 and 152009/18E).

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The data is publically available and the link is provided in Section 3.1.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Anzidei, M.; Argirò, R.; Porfiri, A.; Boni, F.; Anile, M.; Zaccagna, F.; Vitolo, D.; Saba, L.; Napoli, A.; Leonardi, A.; et al. Preliminary clinical experience with a dedicated interventional robotic system for CT-guided biopsies of lung lesions: A comparison with the conventional manual technique. *Eur. Radiol.* **2015**, *25*, 1310–1316. [[CrossRef](#)] [[PubMed](#)]
2. Tekchandani, H.; Verma, S.; Londhe, N.D.; Jain, R.R.; Tiwari, A. Computer aided diagnosis system for cervical lymph nodes in CT images using deep learning. *Biomed. Signal Process. Control* **2022**, *71*, 103158. [[CrossRef](#)]
3. Krilavicius, T.; Zliobaite, I.; Simonavicius, H.; Jaruevicius, L. Predicting respiratory motion for real-time tumour tracking in radiotherapy. In Proceedings of the 2016 IEEE 29th International Symposium on Computer-Based Medical Systems (CBMS), Dublin, Ireland, 20–24 June 2016; pp. 7–12.
4. Brock, K.K.; Mutic, S.; McNutt, T.R.; Li, H.; Kessler, M.L. Use of image registration and fusion algorithms and techniques in radiotherapy: Report of the AAPM Radiation Therapy Committee Task Group No. 132. *Med. Phys.* **2017**, *44*, e43–e76. [[CrossRef](#)] [[PubMed](#)]
5. Han, R.; Jones, C.K.; Lee, J.; Wu, P.; Vagdargi, P.; Uneri, A.; Helm, P.A.; Luciano, M.; Anderson, W.S.; Siewerdsen, J.H. Deformable MR-CT image registration using an unsupervised, dual-channel network for neurosurgical guidance. *Med. Image Anal.* **2022**, *75*, 102292. [[CrossRef](#)]
6. Vishnevskiy, V.; Gass, T.; Szekely, G.; Tanner, C.; Goksel, O. Isotropic total variation regularization of displacements in parametric image registration. *IEEE Trans. Med. Imaging* **2016**, *36*, 385–395. [[CrossRef](#)]
7. Beg, M.F.; Miller, M.I.; Trounev, A.; Younes, L. Computing large deformation metric mappings via geodesic flows of diffeomorphisms. *Int. J. Comput. Vis.* **2005**, *61*, 139–157. [[CrossRef](#)]
8. Cao, Y.; Miller, M.I.; Winslow, R.L.; Younes, L. Large deformation diffeomorphic metric mapping of vector fields. *IEEE Trans. Med. Imaging* **2005**, *24*, 1216–1230.
9. Glaunès, J.; Qiu, A.; Miller, M.I.; Younes, L. Large deformation diffeomorphic metric curve mapping. *Int. J. Comput. Vis.* **2008**, *80*, 317–336. [[CrossRef](#)]
10. Rogelj, P.; Kovačič, S. Symmetric image registration. *Med. Image Anal.* **2006**, *10*, 484–493. [[CrossRef](#)]
11. Beg, M.F.; Khan, A. Symmetric data attachment terms for large deformation image registration. *IEEE Trans. Med. Imaging* **2007**, *26*, 1179–1189. [[CrossRef](#)]
12. Avants, B.B.; Epstein, C.L.; Grossman, M.; Gee, J.C. Symmetric diffeomorphic image registration with cross-correlation: evaluating automated labeling of elderly and neurodegenerative brain. *Med. Image Anal.* **2008**, *12*, 26–41. [[CrossRef](#)] [[PubMed](#)]
13. Balakrishnan, G.; Zhao, A.; Sabuncu, M.R.; Gutttag, J.; Dalca, A.V. VoxelMorph: A learning framework for deformable medical image registration. *IEEE Trans. Med. Imaging* **2019**, *38*, 1788–1800. [[CrossRef](#)]
14. De Vos, B.D.; Berendsen, F.F.; Viergever, M.A.; Sokooti, H.; Staring, M.; Išgum, I. A deep learning framework for unsupervised affine and deformable image registration. *Med. Image Anal.* **2019**, *52*, 128–143. [[CrossRef](#)] [[PubMed](#)]

15. Liu, L.; Aviles-Rivero, A.I.; Schönlieb, C.B. Contrastive Registration for Unsupervised Medical Image Segmentation. *arXiv* **2020**, arXiv:2011.08894.
16. Hering, A.; Häger, S.; Moltz, J.; Lessmann, N.; Heldmann, S.; van Ginneken, B. CNN-based lung CT registration with multiple anatomical constraints. *Med. Image Anal.* **2021**, *72*, 102139. [[CrossRef](#)] [[PubMed](#)]
17. Kim, B.; Kim, D.H.; Park, S.H.; Kim, J.; Lee, J.G.; Ye, J.C. CycleMorph: Cycle consistent unsupervised deformable image registration. *Med. Image Anal.* **2021**, *71*, 102036. [[CrossRef](#)]
18. Liu, L.; Huang, Z.; Liò, P.; Schönlieb, C.B.; Aviles-Rivero, A.I. PC-SwinMorph: Patch Representation for Unsupervised Medical Image Registration and Segmentation. *arXiv* **2022**, arXiv:2203.05684.
19. Zou, J.; Gao, B.; Song, Y.; Qin, J. A review of deep learning-based deformable medical image registration. *Front. Oncol.* **2022**, *12*, 1047215. [[CrossRef](#)]
20. Balakrishnan, G.; Zhao, A.; Sabuncu, M.R.; Guttag, J.; Dalca, A.V. An unsupervised learning model for deformable medical image registration. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Karlstad, Sweden, 18–21 June 2018; pp. 9252–9260.
21. Schreibmann, E.; Chen, G.T.; Xing, L. Image interpolation in 4D CT using a BSpline deformable registration model. *Int. J. Radiat. Oncol. Biol. Phys.* **2006**, *64*, 1537–1550. [[CrossRef](#)]
22. Avants, B.B.; Tustison, N.; Song, G. Advanced normalization tools (ANTs). *Insight J.* **2009**, *2*, 1–35.
23. Marstal, K.; Berendsen, F.; Staring, M.; Klein, S. SimpleElastix: A user-friendly, multi-lingual library for medical image registration. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops, Las Vegas, NV, USA, 27–30 June 2016; pp. 134–142.
24. Modat, M.; McClelland, J.; Ourselin, S. Lung registration using the NiftyReg package. *Med. Image Anal. Clin. Grand Chall.* **2010**, *2010*, 33–42.
25. Rohé, M.M.; Datar, M.; Heimann, T.; Sermesant, M.; Pennec, X. SVF-Net: learning deformable image registration using shape matching. In Proceedings of the International Conference on Medical Image Computing and Computer-Assisted Intervention, Quebec City, QC, Canada, 11–13 September 2017; pp. 266–274.
26. Pei, Y.; Zhang, Y.; Qin, H.; Ma, G.; Guo, Y.; Xu, T.; Zha, H. Non-rigid craniofacial 2D-3D registration using CNN-based regression. In *Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support*; Springer: Berlin/Heidelberg, Germany, 2017; pp. 117–125.
27. Uzunova, H.; Wilms, M.; Handels, H.; Ehrhardt, J. Training CNNs for image registration from few samples with model-based data augmentation. In Proceedings of the International Conference on Medical Image Computing and Computer-Assisted Intervention, Quebec City, QC, Canada, 11–13 September 2017; pp. 223–231.
28. Jaderberg, M.; Simonyan, K.; Zisserman, A. Spatial transformer networks. *Adv. Neural Inf. Process. Syst.* **2015**, *28*, 1–9.
29. Nazib, A.; Fookes, C.; Salvado, O.; Perrin, D. A multiple decoder CNN for inverse consistent 3D image registration. In Proceedings of the 2021 IEEE 18th International Symposium on Biomedical Imaging (ISBI), Nice, France, 13–16 April 2021; pp. 904–907.
30. Mok, T.C.; Chung, A.C. Unsupervised Deformable Image Registration with Absent Correspondences in Pre-operative and Post-recurrence Brain Tumor MRI Scans. In Proceedings of the Medical Image Computing and Computer Assisted Intervention–MICCAI 2022: 25th International Conference, Singapore, 18–22 September 2022; pp. 25–35.
31. Zhang, L.; Ning, G.; Zhou, L.; Liao, H. Symmetric pyramid network for medical image inverse consistent diffeomorphic registration. In *Computerized Medical Imaging and Graphics*; Elsevier: Amsterdam, The Netherlands, 2023; p. 102184.
32. Hu, Y.; Modat, M.; Gibson, E.; Li, W.; Ghavami, N.; Bonmati, E.; Wang, G.; Bandula, S.; Moore, C.M.; Emberton, M.; et al. Weakly-supervised convolutional neural networks for multimodal image registration. *Med. Image Anal.* **2018**, *49*, 1–13. [[CrossRef](#)] [[PubMed](#)]
33. Xu, Z.; Niethammer, M. DeepAtlas: Joint semi-supervised learning of image registration and segmentation. In Proceedings of the International Conference on Medical Image Computing and Computer-Assisted Intervention, Shenzhen, China, 13–17 October 2019; pp. 420–429.
34. Vaswani, A.; Shazeer, N.; Parmar, N.; Uszkoreit, J.; Jones, L.; Gomez, A.N.; Kaiser, Ł.; Polosukhin, I. Attention is all you need. *Adv. Neural Inf. Process. Syst.* **2017**, *30*, 1–11.
35. Chen, J.; He, Y.; Frey, E.C.; Li, Y.; Du, Y. ViT-V-Net: Vision Transformer for Unsupervised Volumetric Medical Image Registration. *arXiv* **2021**, arXiv:2104.06468.
36. Dosovitskiy, A.; Beyer, L.; Kolesnikov, A.; Weissenborn, D.; Zhai, X.; Unterthiner, T.; Dehghani, M.; Minderer, M.; Heigold, G.; Gelly, S.; et al. An image is worth 16 × 16 words: Transformers for image recognition at scale. *arXiv* **2020**, arXiv:2010.11929.
37. Chen, J.; Frey, E.C.; He, Y.; Segars, W.P.; Li, Y.; Du, Y. Transmorph: Transformer for unsupervised medical image registration. *Med. Image Anal.* **2022**, *82*, 102615. [[CrossRef](#)]
38. Liu, Z.; Lin, Y.; Cao, Y.; Hu, H.; Wei, Y.; Zhang, Z.; Lin, S.; Guo, B. Swin transformer: Hierarchical vision transformer using shifted windows. In Proceedings of the IEEE/CVF International Conference on Computer Vision, Montreal, BC, Canada, 11–17 October 2021; pp. 10012–10022.
39. Liu, L.; Wang, H. Multiple Organ Localization in Dual-Modality PET/CT Images Based on Transformer Network with One-to-One Object Query. In Proceedings of the 2022 4th International Conference on Intelligent Medicine and Image Processing, Tianjin China, 18–21 March 2022; pp. 36–42.

40. Ba, J.L.; Kiros, J.R.; Hinton, G.E. Layer normalization. *arXiv* **2016**, arXiv:1607.06450.
41. He, K.; Zhang, X.; Ren, S.; Sun, J. Deep residual learning for image recognition. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, USA, 27–30 June 2016; pp. 770–778.
42. Castillo, R.; Castillo, E.; Guerra, R.; Johnson, V.E.; McPhail, T.; Garg, A.K.; Guerrero, T. A framework for evaluation of deformable image registration spatial accuracy using large landmark point sets. *Phys. Med. Biol.* **2009**, *54*, 1849. [[CrossRef](#)]
43. Çiçek, Ö.; Abdulkadir, A.; Lienkamp, S.S.; Brox, T.; Ronneberger, O. 3D U-Net: Learning dense volumetric segmentation from sparse annotation. In Proceedings of the International Conference on Medical Image Computing and Computer-Assisted Intervention, Athens, Greece, 17–21 October 2016; Springer: Cham, Switzerland; pp. 424–432.
44. Kingma, D.P.; Ba, J. Adam: A method for stochastic optimization. *arXiv* **2014**, arXiv:1412.6980.

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.