ITC 4/54

Information Technology
and Control
Vol. 54 / No. 4/ 2025
pp. 1288-1306
DOI 10.5755/j01.itc.54.4.40153

MS-VMANet: A Multi-Scale VMamba Attention Registration Network for Efficient Assessment of Regional Pulmonary Ventilation Function from 4DCT

Received 2025/01/13 | Accepted after revision 2025/06/26

HOW TO CITE: Bai, M., Liu, X., Meng, Z., Yang, J., Liu, Y. (2025). MS-VMANet: A Multi-Scale VMamba Attention Registration Network for Efficient Assessment of Regional Pulmonary Ventilation Function from 4DCT. *Information Technology and Control,* 54(4), 1288-1306. https://doi.org/10.5755/j01.itc.54.4.40153

MS-VMANet: A Multi-Scale VMamba Attention Registration Network for Efficient Assessment of Regional Pulmonary Ventilation Function from 4DCT

Mengyuan Bai, Xiaofang Liu*, Zijun Meng, Jinfeng Yang, Yang Liu

School of Information Engineering, China Jiliang University, Hangzhou, 310018, China;
e-mails: baimengyuan@cjlu.edu.cn (Mengyuan Bai); liuxfang@cjlu.edu.cn (Xiaofang Liu); 14a0302134@cjlu.edu.cn (Zijun Meng); P23030854053@cjlu.edu.cn (Jinfeng Yang); P23030854030@cjlu.edu.cn (Yang Liu)

Corresponding author: liuxfang@cjlu.edu.cn

The evaluation of regional pulmonary ventilation function is of significant clinical value, particularly in the initial diagnosis of pulmonary disorders, staging assessment, and personalized treatment planning. This study proposes a multi-scale VMamba attention registration network (MS-VMANet) to predict 4DCT pulmonary ventilation changes using unsupervised learning registration. MS-VMANet primarily integrates the efficient visual mamba attention, which captures long-range feature information globally, and the multi-head dilated regional attention improves deformation field prediction via aggregating multi-scale contextual features through dilated convolutions and attention mechanisms. Then, the deformation fields were calculated using the Jacobian determinant to generate images that reflect lung ventilation distribution to assess regional lung ventilation function. According to the experimental findings, the MS-VMANet performs better in terms of registration accuracy and performance, providing a reliable technical means for assessing regional pulmonary ventilation function.

KEYWORDS: 4DCT Lung, Unsupervised Learning, Medical Image Registration, Pulmonary Ventilation

1. Introduction

Lung cancer is currently one of the malignant tumors with the highest incidence and mortality rates worldwide, seriously affecting human health and survival [4]. Radiotherapy has become an important means of treating lung cancer. However, radiotherapy may cause the normal tissues of the human body to be affected by toxic complications, triggering Radiation-induced Lung Injury (RILI) [2], which has a significant impact on patients' lives during or after radiotherapy. The research by Vinogradskiy et al. [44] indicates that during radiotherapy, the risk of RILI can be reduced by selectively avoiding the hyperventilated areas of the lungs. Therefore, accurately assessing pulmonary ventilation function has become a key link in the diagnosis and management of pulmonary disorders. Pulmonary ventilation assessment can comprehensively reflect the lung function status of the patient, which helps reduce the side effects related to radiotherapy and improve the therapeutic effect.

Traditional imaging techniques like MRI, PET, and SPECT can be utilized to measure regional lung ventilation [1]. However, all these methods have inevitable limitations: Firstly, radioactive substances need to be inhaled or injected as tracers, which poses a potential risk of radiation exposure. Secondly, these imaging technologies usually rely on complex equipment and operation procedures and are costly, which limits their application in medical diagnosis and treatment.

With the development of 4DCT [15], using image registration technology to evaluate the regional ventilation function of 4DCT images has emerged as a key method within medical image analysis. This method performs deformable registration on 4DCT scans of the lungs at various breathing stages to obtain a deformable field that describes the spatial transformation relationship of lung pixels. Then, the distribution map of the ventilation function of each area of the lungs is obtained by calculating the Jacobian determinant of the deformation field. Therefore, precise lung image registration plays a critical role in evaluating pulmonary ventilation function.

Traditional lung image registration methods usually regard registration as a process of optimizing the objective function. By solving the optimal solution of spatial geometric transformation, the two images can achieve the best matching in space. Commonly used traditional registration methods such as cubic B-spline [29], elastomer models [3], and diffusion models [42], although they have high registration accuracy, have limited their promotion in real applications because of the significant computational cost and reliance on manual parameter adjustment. The rapid evolution of deep learning technology has brought new breakthroughs to image registration. Neural network models such as VoxelMorph [4], VoxelMorph++ [19], FCN [12], DenseNet [22], and GraphRegNet [16] perform well in the task of lung image registration. However, since these models rely on the estimation of a single deformation field, they are unable to effectively handle large-scale deformations, resulting in a decrease in the registration accuracy of the models. To solve these problems, researchers have begun to adopt multi-stage and multi-scale registration models in Lung image registration, such as RCN [47], RNN [18], LapIPNet [32], DualPRNet [25], mlVIRNET [21], Lung-CRNet [31], PRNet [45], and DefTransNet [33], etc. Although these methods have remarkable effects, they have deficiencies in capturing long-distance spatial dependencies in images, resulting in a decrease in their accuracy when dealing with complex deformations.

Therefore, Transformer is increasingly used in image registration, like TransMorph [6] and TransMatch [7]. In lung image registration, respiratory motion induces substantial anatomical variations in CT images acquired during various breathing stages. Transformer-based methods are highly effective at capturing long-range dependencies. However, they may struggle to effectively extract global contextual information when handling complex lung deformations. Meanwhile, the Transformer usually has a high computational cost, posing additional challenges for real-time applications.

To handle these limitations, this study presents a novel multi-scale VMamba attention registration network (MS-VMANet) for the accurate assessment of regional lung ventilation. MS-VMANet introduces multi-scale feature extraction and hierarchical deformation field optimization strategies, effectively fusing global context information and local detail features, improving the precision of lung registration, and thereby enhancing the accuracy of regional lung ventilation function assessment. The main contributions of this study are summarized as follows:

1 We propose an efficient visual mamba attention (EVMA) module for image feature extraction. EVMA can capture long-distance spatial relationships and global information while reducing the complexity of the model.

2 We propose a multi-head dilated regional attention (MH-DRA) module. MH-DRA can exponentially increase the receptive field and enhance the capture of long-range dependencies, thereby improving the precision of deformation field estimation.

3 We conducted experiments on a publicly available 4DCT dataset of the lungs. According to the experimental data, the MS-VMANet has the highest registration accuracy. Meanwhile, by using the obtained deformation field with higher accuracy, its Jacobian determinant is further calculated to generate a high-precision pulmonary ventilation function map, thereby achieving a more precise evaluation of regional pulmonary function.

2. Related Work

2.1. Calculation of Image Registration-Based Lung Ventilation Imaging

Regional lung ventilation assessment based on image registration is a rapidly developing technique in radiation oncology, with broad application prospects. Currently, the computation of pulmonary ventilation images primarily relies on two main approaches: the Jacobian approach [37, 9] and the Hounsfield Unit (HU) approach [15, 26, 30]. The method based on the Jacobian to obtain the pulmonary ventilation image involves computing the Jacobian determinant of the deformation field via deformable image registration (DIR), without particularly considering the values of the initial CT image. The method based on HU relies on the reconstructed images of various breathing stages and is founded on the linear combination model of lung tissue represented as air and "tissue" components, which infers the changes in lung respiration volume by comparing HU values.

Studies on evaluating pulmonary ventilation function using these two methods have been widely carried out. Reinhardt et al. [37] used the Jacobian determinant of the deformation field obtained through registration for the purpose of measuring regional lung ventilation changes. Ding et al. [9] presented an approach to calculating the ventilation changes in the lung regions. This method matches the maximum expiratory phase of the 4DCT image to the maximum inspiratory phase through the Jacobian determinant, thereby obtaining the ventilation conditions of the lungs in the two stages. Guerrero et al. [15] obtained the ventilation of the lungs by quantifying the changes in voxel density at the two endpoints of the respiratory cycle and using the changes in HU. Kipritidis et al. [26] presented a method for calculating pulmonary ventilation volume by directly scaling the HU value. Experiments show that the resulting ventilation images have significant potential for evaluating the air volume changes in lung regions. Li et al. [30] presented a novel approach to calculating lung ventilation by proportionally combining the recorded HU values with the locally scaled Jacobian determinant, thereby improving the registration accuracy and making it more suitable for describing lung deformation.

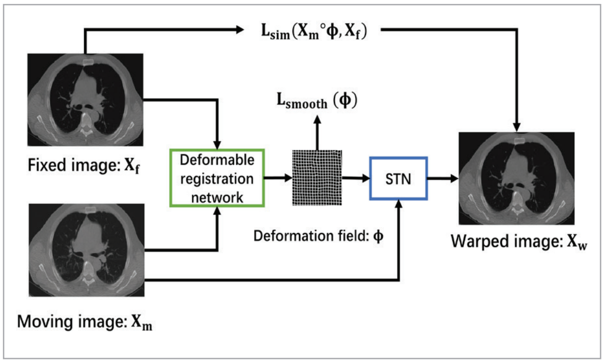
2.2. Lung Image Registration Based on Deep Learning

At present, lung CT registration techniques utilizing deep learning can be classified as supervised registration methods as well as unsupervised registration methods. Supervised registration approaches need genuine deformation fields as labeled data in the training process. In contrast, unsupervised registration methods rely only on image data for training and do not require additional annotation information. Teng et al. [41] utilized a supervised convolutional network for registering adjacent breathing phases, thereby obtaining the corresponding deformation field. Foote et al. [13] designed a patient-specific motor domain method combined with deep convolutional neural networks for accomplishing 2D-3D deformable lung registration. However, the acquisition of deformation fields by these supervised registration methods is usually costly, and at the same time, the quality of the deformation field greatly affects the registration accuracy. Thus, the research focus of image registration has shifted to the unsupervised registration approach that does not rely on the actual deformation field.

The unsupervised registration method updates network parameters by optimizing the dissimilarity of fixed and warped images after spatial transformation. Firstly, by inputting the fixed image X_f and the moving image X_m into the deformable image registration network, the corresponding deformation field ϕ is obtained. Then, ϕ warps X_m via Spatial Transformer Networks (STN) [24] to achieve the deformation processing of the image and thereby obtain the warped image X_w. Finally, the network parameters are iteratively updated by calculating the similarity measure between X_f and X_w. The deformable registration network is deemed optimized when the similarity reaches its maximum. The detailed implementation process is illustrated in Figure 1, where $L_{sim}(X_m \circ \phi, X_f)$ and $L_{smooth}(\phi)$ represent the registration loss functions.

Figure 1

Image registration framework based on unsupervised learning method.


$L_{sim}(X_m{\circ}\phi, X_f)$
$L_{smooth}(\phi)$
Fixed image: X_f
Deformable registration network
STN
Deformation field: ϕ
Warped image: X_w
Moving image: X_m

3. Methodology

3.1. Overview of the MS-VMANet Architecture

We present a multi-scale VMamba attention registration network (MS-VMANet), which consists of two primary components: an efficient visual mamba attention (EVMA) encoder for feature extraction and a multi-head dilated regional attention (MH-DRA) decoder for deformation field generation, as illustrated in Figure 2. Table 1 summarizes the key notations employed throughout this paper.

A five-level shared-weight pyramidal network composed of EVMA modules, with channel numbers of 16, 32, 64, 128, and 256, serves as the encoder. Given the fixed images $X_f \in R^{h\times w\times d}$ and the moving images $X_m \in R^{h\times w\times d}$, the EVMA encoder takes them as inputs and extracts features to generate a series of corresponding image features: F_1, F_2, F_3, F_4, and F_5 for the fixed image, and M_1, M_2, M_3, M_4, and M_5 for the moving image. This process can be expressed as:

$$F_i = f^{i}_{\mathrm{EVMA}}(F_{i-1}), \quad (1)$$

$$M_i = f^{i}_{\mathrm{EVMA}}(M_{i-1}). \quad (2)$$

where i∈{1,2,3,4,5}, and F_i and M_i represent the extracted features of fixed and moving images in the i-th layer of the EVMA module, respectively.

During the decoder stage, the deformation field ϕ_5 is initially generated by feeding F_5 and M_5 into the MH-DRA module, as follows:

$$\phi_5 = f_{\text{MH-DRA}}(F_5, M_5). \quad (3)$$

Subsequently, ϕ_5 is used to warp M_4 via the STN, resulting in the warped feature M'_4, as described below:

$$M'_4 = STN(M_4, \phi_5). \quad (4)$$

The deformation field ϕ_4 is then created by passing F_4 and M'_4 into the MH-DRA module, as illustrated below:

$$\phi_4 = f_{\text{MH-DRA}}(F_4, M'_4). \quad (5)$$

This process is repeated iteratively, repeated for i=3,2,1, as follows:

$$\phi_i = f_{\text{MH-DRA}}(F_i, M'_i), \quad (6)$$

$$M'_i = STN(M_i, \phi_{i+1}). \quad (7)$$

Finally, the STN is used to apply ϕ_1 to X_m, yielding X_w, as expressed below:

$$X_w = STN(X_m, \phi_1). \quad (8)$$

3.2. Efficient Visual Mamba Attention Module

The EVMA module consists of Visual Mamba (VMamba) [48] and a dual-stream spatial-channel attention block (DS-SCAB) [38], as illustrated in Figure 3. The EVMA module enhances the capture of distant spatial relationships without introducing additional parameters or computational overhead, while concurrently emphasizing spatial locations and channel features within the feature maps. This capability allows the traditional VMamba to extract global contextual information, which improves the deformation field estimate and lung image registration task accuracy.

Specifically, the input feature $G \in R^{h \times w \times d}$ is processed through two parallel branches. In the first branch, a linear layer, deep convolution, SiLU function, 2D selective scanning mechanism (2D-SSM), and normalization layer are applied, as described below:

$$G_1 = LNorm(\text{2D-SSM}(SiLU(DWConv(Linear(G))))). \qquad (9)$$

In the second branch, the DS-SCAB is utilized to extract channel and spatial feature information, followed by the SiLU activation function. Specifically, the input feature $G \in R^{h \times w \times d}$ is processed through two shared-weight attention modules to extract channel and spatial features. These features are then fused using two convolutional layers (3×3×3 Conv and 1×1×1 Conv) to produce the output feature, which is passed through the SiLU activation function to generate the ultimate output G_2:

$$\begin{aligned} G_c &= CA(Q_{shared}, K_{shared}, V_{channel}) \\ &= V_{channel} \bullet Soft\max\left(\frac{Q^T_{shared} K_{shared}}{\sqrt{d}}\right), \end{aligned} \qquad (10)$$

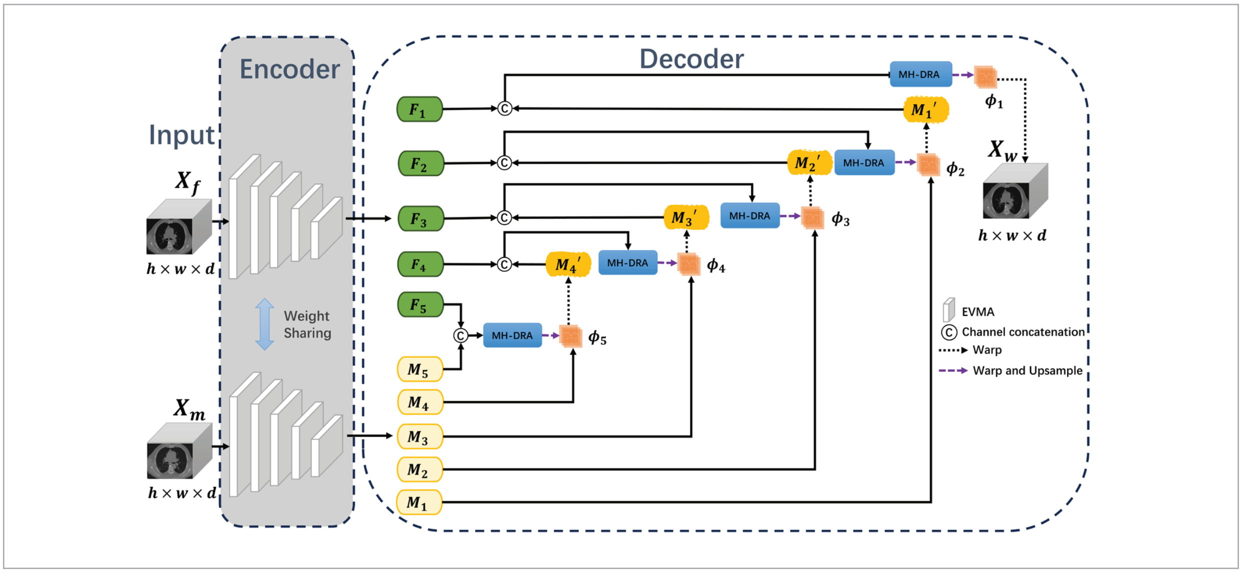
$$\begin{aligned} G_s &= SA(Q_{shared}, K_{shared}, V_{spatial}) \\ &= V_{spatial} \bullet Soft\max\left(\frac{Q_{shared} K^T_{proj}}{\sqrt{d}}\right), \end{aligned} \qquad (11)$$

$$G_2 = SiLU(Conv_{1\times1\times1}(Conv_{3\times3\times3}(G_c + G_s))) \qquad (12)$$

where CA and SA represent the channel attention and spatial attention mechanisms, respectively, while G_c and G_s are the corresponding attention maps. $Q_{shared} = W^Q G$, $K_{shared} = W^K G$, and $V_{shared} = W^V G$ represent the shared query, key, and value vectors, with W^Q, W^K, and W^V being their respective projection weights. $V_{channel} = W^V G$ and $V_3 = W^V G$ represent the channel value layers and space value layers, respectively, each with a vector size of d.

Figure 2

Proposed MS-VMANet architecture. (X_f, X_m, and X_w represent the fixed image, the moving image, and the warped image, respectively. F_5, F_4, F_3, F_2, F_1, and M_5, M_4, M_3, M_2, M_1 represent the features extracted from the fixed images and the moving images at each level, respectively. M'_4, M'_3, M'_2, and M'_1 denote the warped moving image features at each level. ϕ_5, ϕ_4, ϕ_3, ϕ_2, ϕ_1 are the deformation fields obtained at each level.)



Then, the outputs from the two branches are combined using the Hadamard product to generate the final output G_{out}, as illustrated below:

$$G_{out} = Linear(G_1 \odot G_2)\,. \tag{13}$$

3.3. Multi-Head Dilated Regional Attention Module

Generating an accurate deformation field is essential for lung image registration. It is essential for ensuring precise alignment associated with fixed and moving images. Thus, we employ the multi-head dilated regional attention module (MH-DRA) [17], as illustrated in Figure 4. The MH-DRA module expands the regional attention mechanism of neighborhood attention to sparse global attention with reduced restrictions. This approach captures a broader global context while significantly expanding the receptive field, thereby enabling more precise generation of the deformation field.

Table 1

Notations and their definitions.

Notation	Definition	Notation	Definition
X_f	The fixed image	$V_{spatial}$	The spatial attention value vector
X_m	The moving image	m	Number of attention heads in the MH-DRA
X_w	The warped image	δ	The dilation factor of the MH-DRA
ϕ_i	The deformation field in the i-th layer	g(x)	The k×k×k neighborhood of voxel x
L_{sim} (*)	The spatial regularization term of ϕ	B	The learnable relative position bias
L_{smooth} (*)	The similarity measurement term of ϕ	λ	The regularization parameter of the loss function
h	The height of the input feature	pi	The voxel of index i
w	The width of the input feature	$\widehat{X_f}$	X_f with the local average pixel value subtracted
d	The depth of the input feature	$\widehat{X_m}$	X_m with the local average pixel value subtracted
c	The number of channels of the input feature	$U_x(A)$	The estimated deformation field components in the x direction at point A (x, y, z)
F_i	The features of X_f in the i-th layer	$U_y(A)$	The estimated deformation field components in the y direction at point A (x, y, z)
M_i	The features of X_m in the i-th layer	$U_z(A)$	The estimated deformation components in the z direction at point A (x, y, z)
M_i’	The processed features of X_m in the i-th layer	u(*)	Displacements in the direction of x
G	The input feature of EVMA	v(*)	Displacements in the direction of y
G_c	The channel attention map of EVMA	w(*)	Displacements in the direction of z
G_s	The spatial attention map of EVMA	S_{X_f}	The regions of interest of X_f
Q	The query vector	S_{X_w}	The regions of interest of X_w
K	The key vector	μ_{X_f}	The mean values of X_f
V	The value vector	μ_{X_w}	The mean values of X_w
W^Q	The projection weight matrix of Q	σ_{X_f}	The standard deviations of X_f
W^K	The projection weight matrix of K	σ_{X_w}	The standard deviations of X_w
W^V	The projection weight matrix of V	c1	A constant
$V_{channel}$	The channel attention value vector	c2	A constant

Given the inputs fixed image features $F \in R^{h \times w \times d}$ and moving image features $M \in R^{h \times w \times d}$, the corresponding query vector Q and key vector K are obtained through linear projection and layer normalization, as illustrated below:

$$Q = LNorm(L\Pr oj(F)), \tag{14}$$

$$K = LNorm(L\Pr oj(M)), \tag{15}$$

where $Q, K \in R^{h \times w \times d \times c/m}$, with h, w, d, and c denoting the height, width, depth, and number of channels of the input features, respectively. m represents the count of attention heads in MH-DRA.

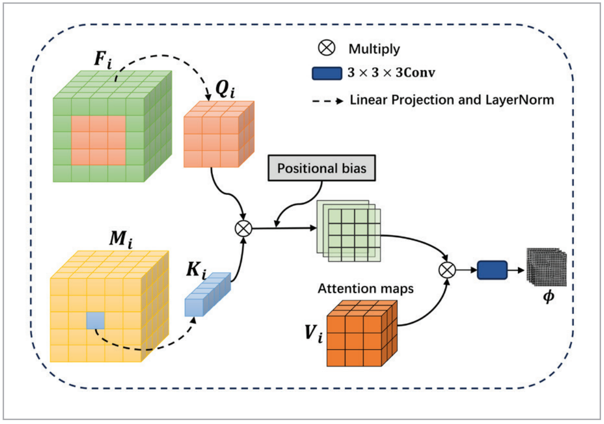
Next, the m-th attention head's attention map can be computed using the formula below:

$$AM_{\delta}^{m} = Soft\max(Q_{x}^{m} \cdot (K_{g^{\delta}(x)}^{m})^{T} + B^{m}(x, g^{\delta}(x))), \tag{16}$$

where δ denotes the dilation factor, g(x) represents the k×k×k neighborhood of voxel x, and $B \in R^{m \times k \times k \times k}$ represents a learnable relative position bias.

Figure 4

The structure of the MH-DRA module.



Then, the deformation field is weighted using AM to generate a series of sub-deformation fields. These sub-deformation fields, computed from each attention head, are integrated through a 3D convolution layer (3×3×3 Conv) to generate the deformation field ϕ_i for the i-th level, as illustrated below:

$$\phi^{m} = AM_{\delta}^{m} \cdot V_{\delta}, \tag{17}$$

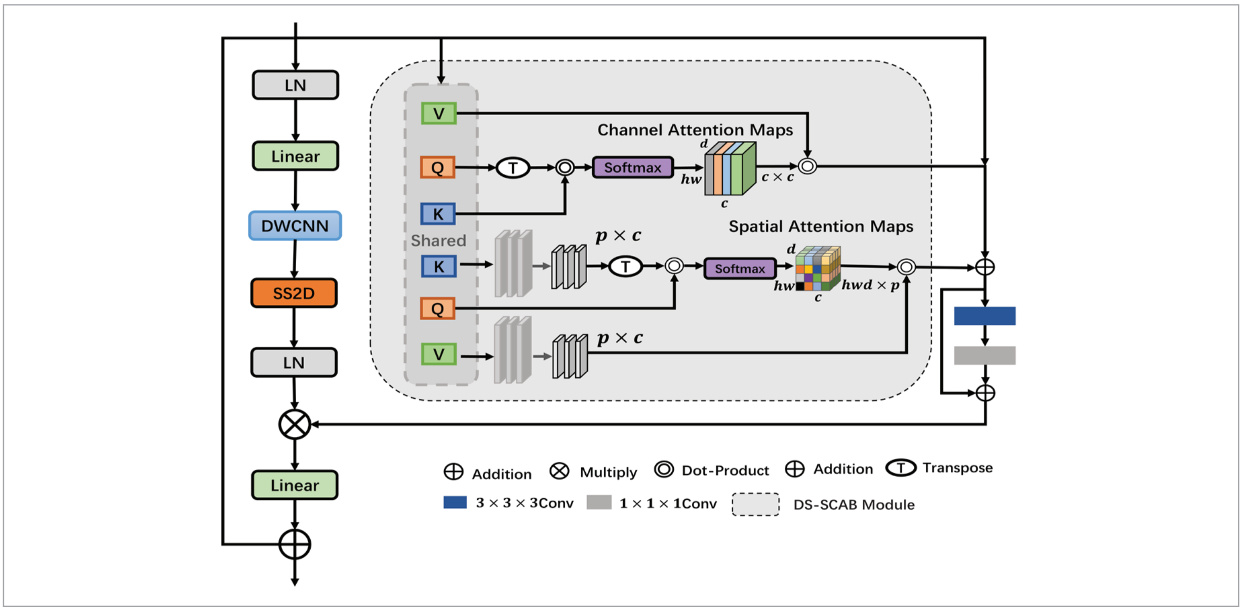
$$\phi_{i} = Conv_{3\times3\times3}(Concat(\phi_{i}^{1}, \phi_{i}^{2}, \phi_{i}^{3}, ..., \phi_{i}^{m-1}, \phi_{i}^{m})). \tag{18}$$

3.4. Loss Lunction

In this experiment, the overall loss function L_{total} is composed of two components: L_{smooth} enforces spatial

Figure 3

Schematic diagram of EVMA module structures.



regularization on the deformation field produced by the registration network, ensuring its smoothness. Meanwhile, L_{sim} evaluated the similarity among the fixed and warped images. The definition of L_{sim} is as follows:

$$L_{sim}(X_f, X_m, \phi) = -\sum_{p \in \Omega} \frac{\sum p_i((X_f(p_i) - \widehat{X_f(p)})(X_m(\phi(p_i)) - \widehat{X_m(\phi(p))}))^2}{\sum p_i(X_f(p_i) - \widehat{X_f(p)})^2 (X_m(\phi(p_i)) - \widehat{X_m(\phi(p))})^2}, \quad (19)$$

where X_f and X_m refer to the fixed and moving images, respectively. ϕ represents the deformation field, pi represents a voxel with index i, $\widehat{X_f}$ denotes the fixed image with the local average pixel value subtracted, and $\widehat{X_m}$ denotes the moving image with the local average pixel value subtracted.

During network training, discontinuous deformation fields are often generated as a result of optimizing the image similarity measure. To prevent overlapping in the predicted deformation fields, a spatial smoothness constraint is typically incorporated. The formula is as follows:

$$L_{smooth}(\phi) = \sum_{p \in \Omega} \left\| \nabla \phi(p) \right\|^2, \quad (20)$$

where $\nabla\phi(p)$ represents the spatial gradient of voxel p.

The formula for the total loss function L_{total} includes a weighted spatial regularisation term and a similarity measurement term:

$$L_{\text{total}} = L_{sim}(X_f, X_m, \phi) + \lambda L_{smooth}(\phi), \quad (21)$$

where λ denotes the regularization parameter.

3.5. Jacobian Determinant

The Jacobian determinant evaluates the regional variance in volume between two images by calculating the derivatives of the displacement [34]. The formula is as follows:

$$Jac(A) = \begin{vmatrix} 1 + \frac{\partial U_x(A)}{\partial x} & \frac{\partial U_x(A)}{\partial y} & \frac{\partial U_x(A)}{\partial z} \\ \frac{\partial U_y(A)}{\partial x} & 1 + \frac{\partial U_y(A)}{\partial y} & \frac{\partial U_y(A)}{\partial z} \\ \frac{\partial U_z(A)}{\partial x} & \frac{\partial U_z(A)}{\partial y} & 1 + \frac{\partial U_z(A)}{\partial z} \end{vmatrix}, \quad (22)$$

where U_x (A), U_y (A), and U_z (A) respectively represent the estimated deformation components along the x, y, and z directions at point A (x, y, z). If Jac = 0, it implies the absence of expansion or contraction in the lung region. When 0 < Jac < 1, it indicates the presence of lung region expansion. If Jac < 0, it signifies lung region shrinkage [40].

4. Experiments

4.1. Dataset

As shown in Table 2, all experiments were conducted using three publicly available lung 4DCT datasets: (1) the 4D-Lung dataset [42]; (2) the CREATIS dataset [43]; (3) the DIRLAB dataset [5]. Each scan contains 10 images corresponding to the 10 phases of the respiratory cycle. Among these, landmarks at the maximum inspiratory and maximum expiratory phases have been manually identified in the CREATIS and DIRLAB datasets, which can be used to evaluate registration performance.

Table 2

Detailed information on the 4D-Lung, CREATIS, and DIRLAB datasets.

Dataset	4D-Lung	CREATIS	DIRLAB
Patients	20	6	10
Modality	4DCT	4DCT	4DCT
Format	.dicom	.dicom	.img
Landmarks	-	100	300
Function	Train	Train	Test

The training phase utilized the 4D-Lung dataset, which included 18 scans selected at random for training and 2 for validation, as well as the CREATIS dataset, consisting of 4 scans chosen at random for training and 2 for validation. During the training process, a leave-one-out cross-validation method was adopted, and the end-inhalation and end-exhalation images from the DIRLAB dataset were used for testing.

Since the tissue structure in the original lung images is not clear enough, it is necessary to enhance the brightness and contrast and normalize the image intensity to the range of [0,1] to improve the image quality. Then, a clustering algorithm is applied to eliminate bed plates and background noise in the image to the clarity of the lung region. Finally, all 4DCT lung images were cropped to 192 × 192 × 192 and resampled to a consistent voxel size of 1 mm × 1 mm × 1 mm.

4.2. Implementation Details

All experiments were performed on an NVIDIA GeForce RTX 4090 utilizing the PyTorch deep learning framework [34]. During training, the attention heads of MH-DRA were set to 8, 4, 2, 1, and 1, respectively. The expansion coefficient δ was defined as 4, while the regularization parameter λ for the total loss was fixed at 0.1. In the experiment, the ADAM optimizer was used, and the learning rate was set to 0.0001. Each training session is set to 500 times, with a batch size of 1.

4.3. Evaluation Metrics

In this study, Target Registration Error (TRE) [35], Dice Similarity Coefficient(DSC) [46], and Structure Similarity Index Measure(SSIM) [10] measures were employed as quantitative metrics to assess registration performance rigorously.

4.3.1. Target Registration Error

TRE is the Euclidean distance between the landmark in the moving image and its matching position in the fixed image. A smaller value of TRE signifies greater registration accuracy of the algorithm. The specific mathematical expression is as follows:

$$TRE = \sqrt{(x+u(x)-x^{'})^2+(y+v(y)-y^{'})^2+(z+w(x)-z^{'})^2}\ , \quad (23)$$

where (x, y, z) denotes the landmark in the fixed image′s landmark, and (x′, y′, z′) denotes the corresponding landmark in the warped image. u(x), v(x), and w(x) represent the displacements along the x, y, and z directions, respectively, following the prediction of the deformation field.

4.3.2. Dice Similarity Coefficient

DSC is employed to assess how similar two images are to each other. Its value spans from 0 to 1, where 0 denotes complete misalignment of the two images, and 1 indicates perfect alignment. The specific mathematical expression is as follows:

$$Dice = \frac{2\,|\,S_{X_f} \cap S_{X_w}\,|}{|\,S_{X_f}\,| + |\,S_{X_w}\,|}, \quad (24)$$

where S_{X_f} and S_{X_w} denote the regions of interest of the two images, respectively.

4.3.3. Structure Similarity Index Measure

SSIM measures the structural similarity of two images. The SSIM value varies between 0 and 1, where values approaching 1 signify greater similarity. The formula is given as follows:

$$SSIM(X_f, \mathrm{X}_w) = \frac{(2\mu_{X_f}\mu_{X_w} + \mathrm{c}_1)(2\sigma_{X_f}\sigma_{X_w} + \mathrm{c}_2)}{(\mu_{X_f}^2 + \mu_{X_w}^2 + \mathrm{c}_1)(\mu_{X_f}^2 + \mu_{X_w}^2 + \mathrm{c}_2)}, \quad (25)$$

where X_f and X_w represent the fixed and warped images, respectively. μ_{X_f} and μ_{X_w} represent the respective mean values of X_f and X_w, σ_{X_f} and σ_{X_w} denote the standard deviations of X_f and X_w. c_1 and c_2 are constants.

4.4. Comparative Experiments Analysis

To prove the efficiency of MS-VMANet for lung registration, the MS-VMANet algorithm was compared with other unsupervised learning registration methods, including VoxelMorph [4], LungRegNet [14], ProgNet [11], and HPRN [23]. The primary experimental findings are given in Tables 3-4.

Table 3 shows the TRE values for different algorithms, while Figure 5 shows the bar chart of TRE. As shown in Table 3 and Figure 5, MS-VMANet significantly reduces registration errors across the 10 cases in the DIRLAB dataset, with an average TRE value approximately 7 percentage points lower than before registration. Furthermore, compared to other methods, MS-VMANet achieves an average target registration error of around 1.78mm on the DIRLAB dataset, outperforming existing registration models and effectively improving the accuracy of lung image registration. However, in instances involving substantial deformations, like Case 7 and Case 8, the registration performance of MS-VMANet is less effective compared to LungRegNet. We speculate that this discrepancy may arise from the MH-DRA module's use of a fixed dilation factor when predicting the deformation field, which may not adequately accommodate the non-uniform nature of lung deformations, leading to insufficient deformation information extraction in certain lung regions. Additionally, before predicting the deformation field, LungRegNet incorporates lung vessel-enhanced images, introducing anatomical structural constraints that improve registration accuracy in vascular regions. Nevertheless, compared to VoxelMorph, ProgNet, and HPRN none of which integrate additional anatomical information MS-VMANet achieves the lowest TRE value and delivers the best registration performance in this study.

Figure 5

Bar chart of TRE values for various registration algorithms.

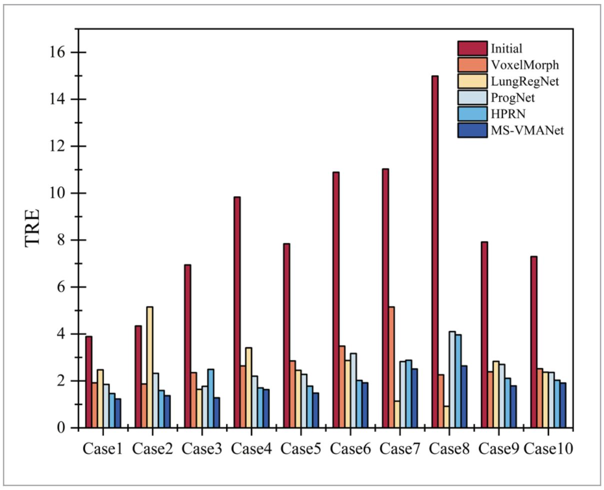


Table 4 presents the DSC and SSIM values for each method. Table 4 shows that MS-VMANet demonstrates significantly superior DSC and SSIM values on the DIRLAB dataset compared to other unsupervised learning methods. MS-VMANet exhibits high average values for DSC (0.904) and SSIM (0.873) in all test cases. Additionally, as shown in Figure 6, the boxplot for the MS-VMANet method is the narrowest, indicating that this method exhibits higher stability and robustness. Furthermore, based on

Figure 6

Box plot of DSC and SSIM values for various registration algorithms.

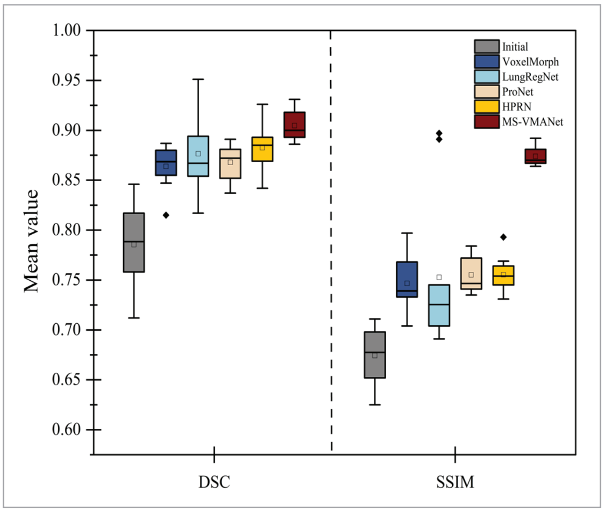


Table 3

A comparison of the registration performance of MS-VMANet with alternative unsupervised learning approaches based on TRE (mm) on the DIRLAB datasets. (Lower TRE values indicate better registration performance.)

Dataset	Initial	VoxelMorph	LungRegNet	ProNet	HPRN	MS-VMANet
Case1	3.89(2.78)	1.92(0.80)	2.47(0.83)	1.85(0.79)	1.46(0.71)	**1.23(0.86)**
Case2	4.34(3.90)	1.87(0.77)	5.15(3.75)	2.32(0.85)	1.59(0.76)	**1.37(0.93)**
Case3	6.94(4.05)	2.35(1.02)	1.64(0.79)	1.77(0.97)	2.49(0.61)	**1.28(0.71)**
Case4	9.83(4.85)	2.64(1.85)	3.41(4.36)	2.20(1.18)	1.70(0.68)	**1.63(1.19)**
Case5	7.84(5.50)	2.85(2.13)	2.45(0.85)	2.28(1.57)	1.78(1.91)	**1.48(0.65)**
Case6	10.89(6.96)	3.48(2.75)	2.87(2.58)	3.17(1.82)	2.02(1.41)	**1.92(0.84)**
Case7	11.03(7.42)	5.15(4.20)	**1.14(0.20)**	2.82(1.72)	2.88(2.71)	2.51(1.78)
Case8	14.99(9.00)	2.26(1.56)	**0.92(0.47)**	4.10(3.18)	3.96(2.89)	2.64(1.39)
Case9	7.92(3.97)	2.39(2.36)	2.83(0.71)	2.70(1.46)	2.11(1.13)	**1.79(0.72)**
Case10	7.30(6.34)	2.52(2.43)	2.37(0.99)	2.36(1.62)	2.03(1.44)	**1.91(0.98)**
Mean and Std	8.46(5.48)	2.74(1.99)	2.53(1.55)	2.56(1.52)	2.20(1.43)	**1.78(1.01)**

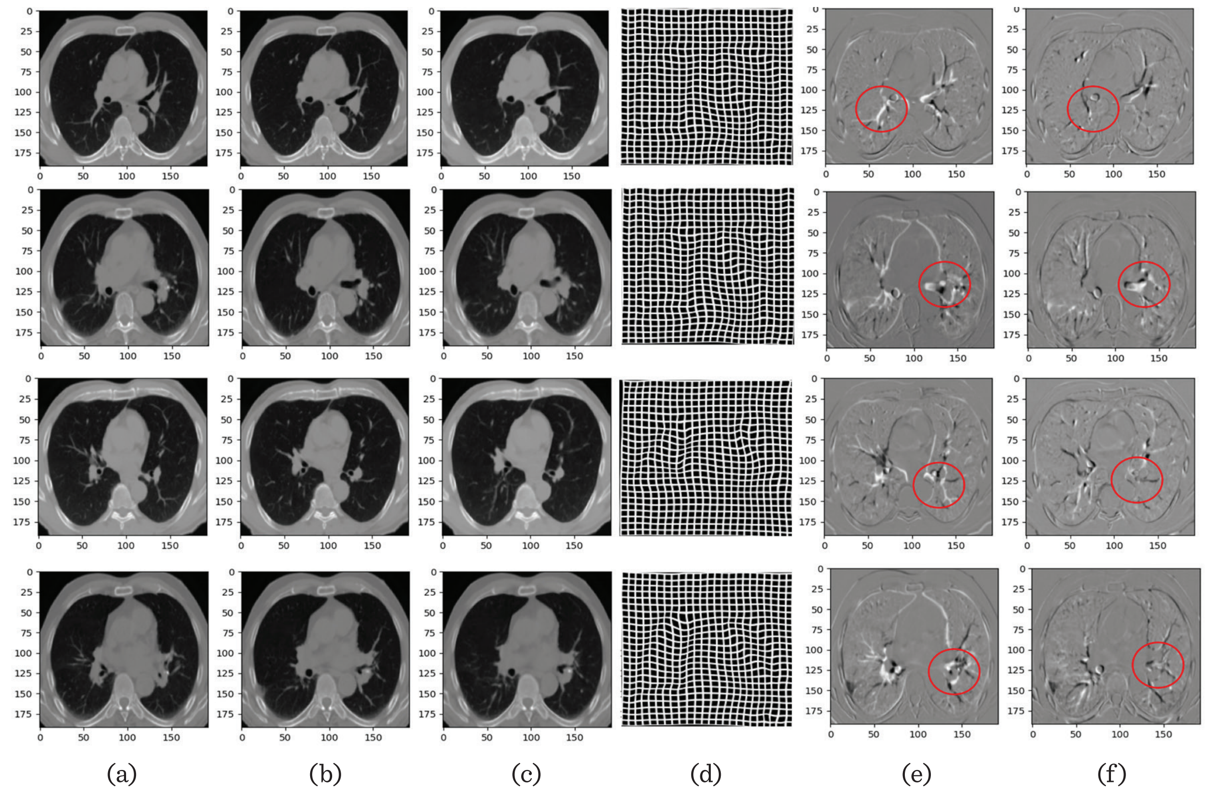
Table 4

A comparison of the registration performance of MS-VMANet with alternative unsupervised learning approaches based on DSC and SSIM on the DIRLAB datasets. (Higher values indicate better registration performance.)

Dataset	Initial		VoxelMorph		LungRegNet		ProNet		HPRN		MS-VMANet	
	DSC	SSIM	DSC	SSIM	DSC	SSIM	DSC	SSIM	DSC	SSIM	DSC	SSIM
Case1	0.846	0.652	0.880	0.782	0.865	0.745	0.891	0.745	0.926	0.769	**0.931**	**0.886**
Case2	0.834	0.646	0.887	0.797	0.817	0.736	0.879	0.774	0.906	0.758	**0.918**	**0.875**
Case3	0.817	0.711	0.873	0.768	0.894	0.701	0.890	0.743	0.869	0.745	**0.897**	**0.881**
Case4	0.781	0.625	0.862	0.741	0.857	0.704	0.872	0.770	0.893	0.750	**0.924**	**0.866**
Case5	0.796	0.707	0.855	0.736	0.869	0.715	0.881	0.735	0.889	0.747	**0.906**	**0.871**
Case6	0.758	0.684	0.847	0.733	0.854	0.709	0.849	0.741	0.887	0.764	**0.893**	**0.868**
Case7	0.733	0.698	0.815	0.704	**0.934**	**0.891**	0.857	0.739	0.850	0.761	0.898	0.864
Case8	0.712	0.665	0.883	0.737	**0.951**	**0.897**	0.837	0.748	0.842	0.731	0.886	0.867
Case9	0.785	0.683	0.871	0.723	0.848	0.739	0.852	0.784	0.883	0.793	**0.893**	**0.892**
Case10	0.792	0.672	0.866	0.745	0.877	0.691	0.872	0.772	0.881	0.735	**0.902**	**0.869**
Mean	0.785	0.674	0.863	0.746	0.876	0.752	0.868	0.751	0.882	0.755	**0.904**	**0.873**

Figure 7

Registration results of the MS-VMANet method on the DIRLAB dataset. (a) Moving image; (b) Fixed image; (c) Warped image; (d) Deformation mesh; (e) Difference image before registration; (f) Difference image after registration.



the median values, the DSC and SSIM values of the proposed MS-VMANet method are the highest, further demonstrating its superiority in lung registration precision.

As shown in Figure 7, significant morphological and volumetric differences are observed when comparing the moving and fixed images. MS-VMANet effectively matches the moving image to the fixed image, minimizing intensity discrepancies among the warped and fixed images while maintaining local texture details in the fixed image. Figure 8 contrasts MS-VMANet with the warped images of the other four unsupervised learning-based methods: VoxelMorph, LungRegNet, ProgNet, and HPRN, to more clearly illustrate the MS-VMANet method's efficacy in lung image registration.

As shown in Figure 8, the MS-VMANet method effectively performs spatial transformations on the moving image's external contours and internal tissue textures to address differences with the fixed image, producing a registration result that closely matches the internal structures and texture features of the fixed image. In contrast, the VoxelMorph, LungRegNet, ProgNet, and HPRN methods exhibit limitations in handling local registration issues, failing to accurately register and deform the moving image to match the internal structures and texture features of the fixed image.

4.5. Ablation Experiment Analysis

Aimed at evaluating the EVMA and MH-DRA's influence on model performance, we substituted them with 3D convolutional blocks that included 3×3×3 convolutions, BatchNorm, and ReLU functions. Tables 5-6 present the ablation study results for various registration approaches on the DIRLAB dataset.

Figure 8

The comparison of registration performance between the MS-VMANet method and other methods on the DIRLAB dataset. (a) Moving image; (b) Fixed image; (c)Warped images obtained by the VoxelMorph method; (d)Warped images obtained by the LungRegNet method; (e)Warped images obtained by the ProgNet method; (f)Warped images obtained by the HPRN method; (g)Warped images obtained by the MS-VMANet method.

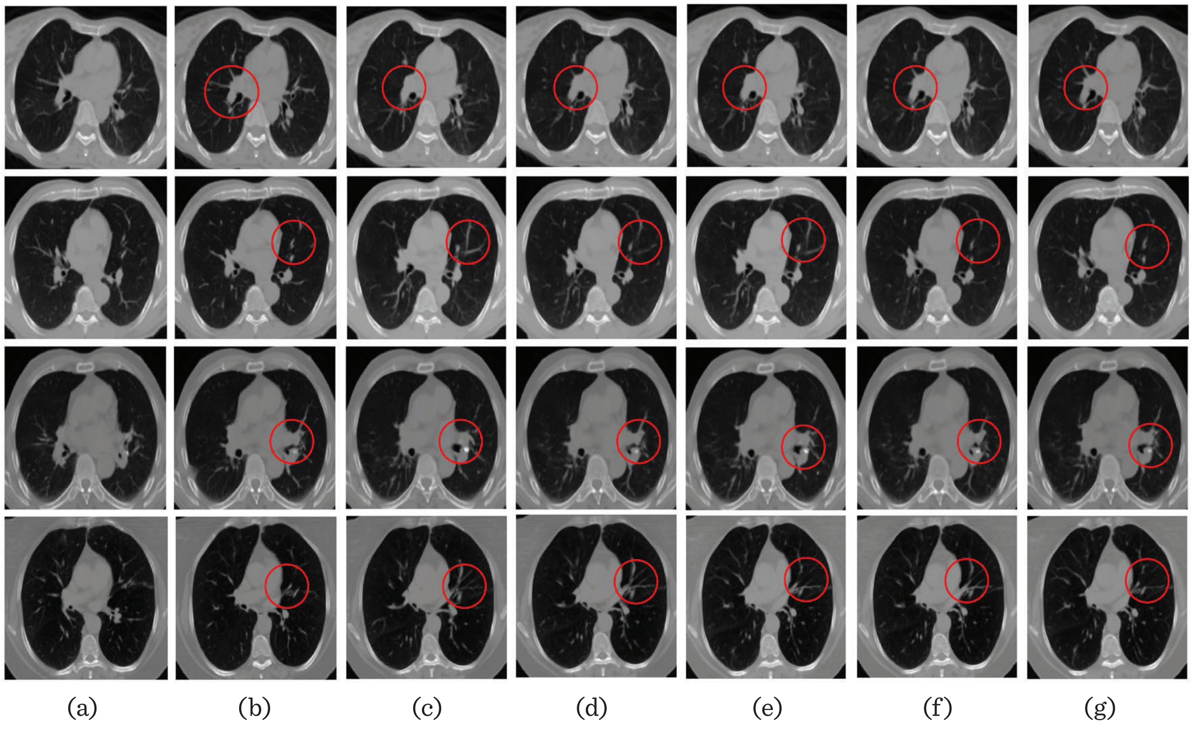


Table 5

TRE performances of ablation studies on the DIRLAB dataset.

Dataset	Initial	MS-VMANet(without EVMA)	MS-VMANet(without MH-DRA)	MS-VMANet
Case1	3.89(2.78)	1.82(0.94)	2.36(1.51)	**1.23(0.86)**
Case2	4.34(3.90)	2.03(1.56)	2.11(1.37)	**1.37(0.93)**
Case3	6.94(4.05)	1.97(1.74)	1.89(0.95)	**1.28(0.71)**
Case4	9.83(4.85)	2.23(1.62)	2.46(1.06)	**1.63(1.19)**
Case5	7.84(5.50)	2.56(1.05)	2.24(1.33)	**1.48(0.65)**
Case6	10.89(6.96)	2.91(1.17)	2.86(1.48)	**1.92(0.84)**
Case7	11.03(7.42)	3.35(2.02)	3.16(1.94)	**2.51(1.78)**
Case8	14.99(9.00)	2.28(1.79)	2.75(1.21)	**2.64(1.39)**
Case9	7.92(3.97)	2.14(1.19)	2.37(1.82)	**1.79(0.72)**
Case10	7.30(6.34)	2.41(1.58)	2.23(1.69)	**1.91(0.98)**
Mean and Std	8.46(5.48)	2.37(1.52)	2.44(1.44)	**1.78(1.01)**

Table 6

DSC and SSIM performances of ablation studies on the DIRLAB dataset.

Dataset	Initial		MS-VMANet (without EVMA)		MS-VMANet (without MH-DRA)		MS-VMANet	
	DSC	SSIM	DSC	SSIM	DSC	SSIM	DSC	SSIM
Case1	0.846	0.652	0.886	0.842	0.891	0.812	**0.931**	**0.886**
Case2	0.834	0.646	0.905	0.853	0.883	0.806	**0.918**	**0.875**
Case3	0.817	0.711	0.892	0.837	0.875	0.785	**0.897**	**0.881**
Case4	0.781	0.625	0.874	0.797	0.862	0.792	**0.924**	**0.866**
Case5	0.796	0.707	0.868	0.821	0.871	0.768	**0.906**	**0.871**
Case6	0.758	0.684	0.852	0.834	0.848	0.761	**0.893**	**0.868**
Case7	0.733	0.698	0.885	0.866	0.839	0.774	**0.898**	**0.864**
Case8	0.712	0.665	0.873	0.838	0.825	0.852	**0.886**	**0.867**
Case9	0.785	0.683	0.912	0.809	0.861	0.781	**0.893**	**0.892**
Case10	0.792	0.672	0.887	0.855	0.864	0.776	**0.902**	**0.869**
Mean	0.785	0.674	0.883	0.835	0.861	0.79	**0.904**	**0.873**

As displayed in Table 5 and Figure 9, eliminating either the EVMA or MH-DRA module leads to a notable rise in TRE values, consequently reducing the accuracy of the image alignment.

Additionally, as indicated in Table 6 and Figure 10, compared to the MS-VMANet without the EVMA or MH-DRA modules, incorporating these modules yields higher DSC and SSIM values, further demonstrating the critical role of the EVMA and MH-DRA modules in enhancing the performance of the MS-VMANet model.

Figure 9

Bar chart of the TRE values for the ablation experiment.

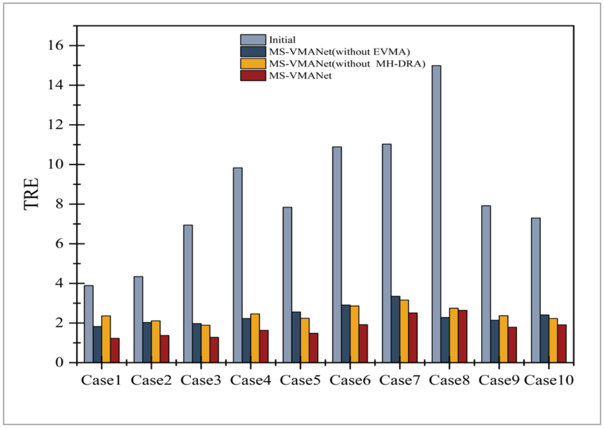
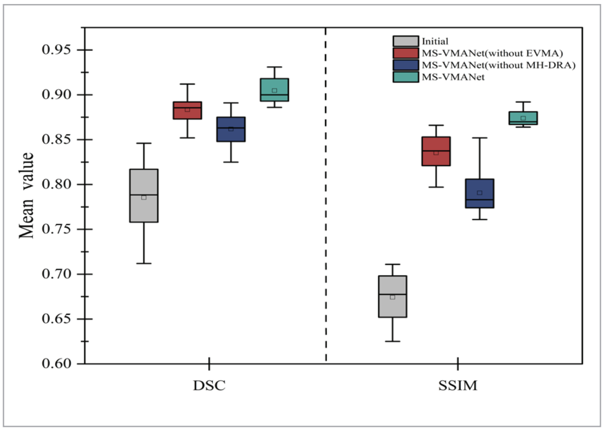


Figure 10

Box plot of DSC and SSIM values for the ablation experiment.



4.6. Complexity Analysis

The time and space complexity of the five methods, including MS-VMANet, were evaluated to comprehensively analyze their computational cost, as shown in Table 7. Here, n=h×w×d denotes the input feature volume's voxel count, and c denotes the dimensionality of feature channels.

Table 7

Comparison of the performance parameters of networks for registration.

Methods	Params	FLOPs	Time Complexity
Voxelmorph	6.3M	84.8G	O(nc)
LungRegNet	14.6M	157.3G	O(nc^2)
ProgNet	5.8M	73.4G	O(nlog^{2}n)
HPRN	7.6M	91.7G	O(nclogc)
MS-VMANet	**2.1M**	**56.5G**	**O(nlogn)**

As presented in Table 7, MS-VMANet demonstrates a considerably reduced computational expense compared to other methods. With only 2.1M parameters and 56.5G FLOPs, MS-VMANet demonstrates superior memory usage and computation efficiency. Moreover, the time complexity of MS-VMANet is O(nlogn), which is significantly more scalable compared to other networks such as LungRegNet with O(nc2) and ProgNet with O(nlog2n). This reduced complexity makes MS-VMANet highly suitable for practical deformable image registration tasks.

4.7. Regional Pulmonary Ventilation Function Analysis

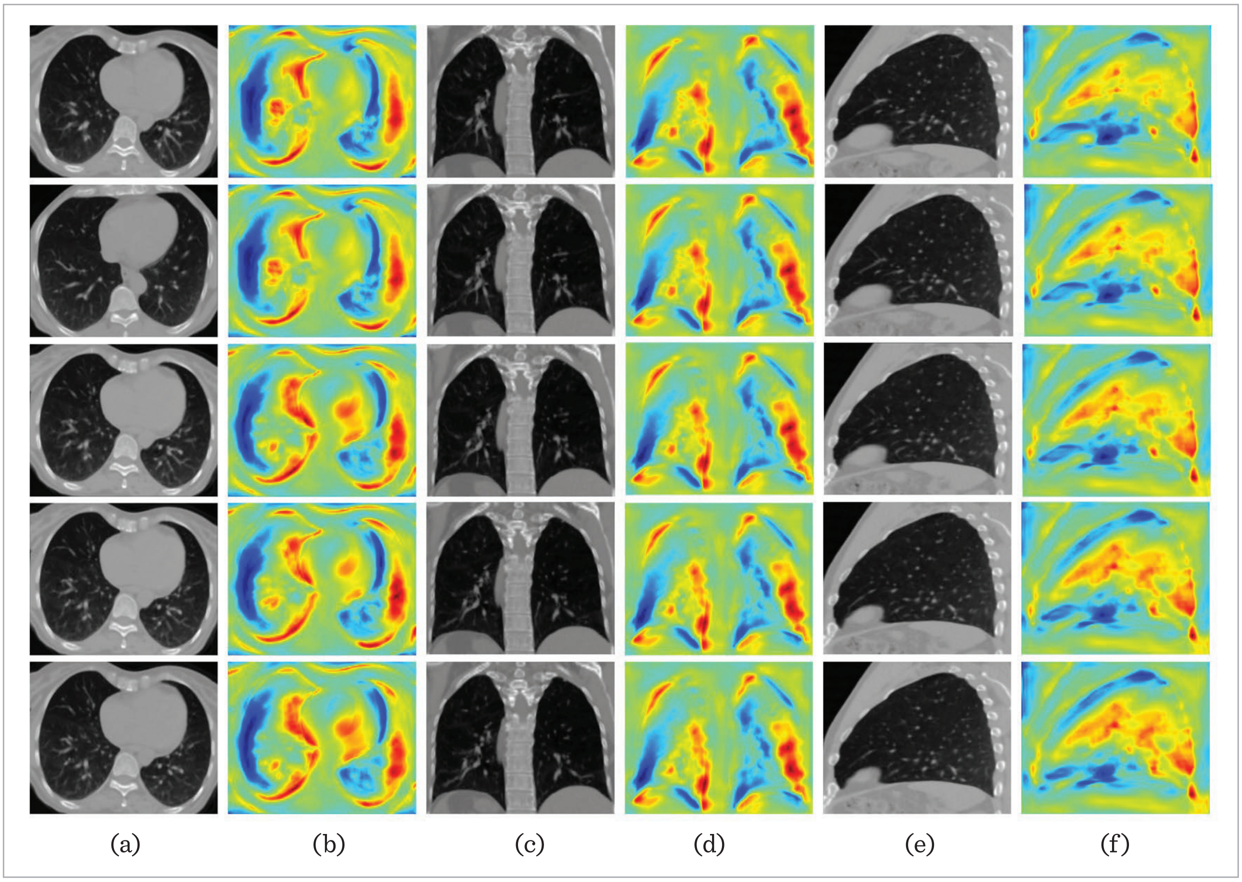
Based on the registration results mentioned above, the effectiveness and accuracy of the MS-VMANet method are validated, demonstrating its significant improvement in registration accuracy while maintaining stability. We further compute the deformation fields of the DIRLAB dataset using the Jacobian determinant to generate regional lung ventilation function images. These images are then converted into pseudo-color representations to visualize and assess the patient's lung ventilation function.

Figure 11 presents the pulmonary ventilation distribution across five consecutive slices in the axial, sagittal, and coronal planes of the DIRLAB dataset. A deeper red color indicates a greater degree of lung voxel expansion, signifying stronger lung ventilation function, whereas a darker blue color represents increased lung voxel contraction, suggesting a corresponding reduction in ventilation function.

As illustrated in Figure 11, the peripheral areas of the lung appear in blue or red, signifying a robust ventilation function in these regions. The central regions of the lung predominantly display green, indicating that these areas neither expand nor contract, which suggests weak or potentially absent ventilation in certain localized regions. Furthermore, pulmonary ventilation is unevenly distributed, with variations observed across different layers. However, a degree of correlation and continuity in ventilation is maintained between consecutive and adjacent layers.

Figure 11

Regional pulmonary ventilation distribution images across five consecutive lung slices in the axial, coronal, and sagittal planes were obtained using MS-VMANet. (a) Lung CT image in the axial plane; (b) Regional pulmonary ventilation distribution image in axial slices; (c) Lung CT image in the coronal plane; (d) Regional pulmonary ventilation distribution image in coronal slices; (e) Lung CT image in the sagittal plane; (f) Regional pulmonary ventilation distribution image in sagittal slices.



Overall, the right lung performs better at ventilation than the left lung, as the right lung exhibits a higher prevalence of blue and red areas. Consequently, in the context of image-guided lung radiation therapy, regions with high ventilation (depicted by the red and blue areas) can be selectively preserved, and the irradiation dose to these areas can be reduced, thereby minimizing potential damage to healthy lung tissue and enhancing therapeutic efficacy.

5. Discussion

Based on deformable image registration, we presented the MS-VMANet approach in this study for evaluating lung ventilation function. While the experimental results validate the efficacy of MS-VMANet, it is essential to recognize the limitations and possible avenues for improvement in the proposed method.

1 Registration of Large Deformations

For lung CT images with significant deformations, although MS-VMANet can effectively handle these cases, its performance still lags behind that of LungRegNet. This can be attributed to two main factors: First, MS-VMANet employs a fixed dilation factor as proposed in [17] when generating the deformation field, restricting the model's capacity to adjust to the non-uniformity of lung deformations, thereby reducing the accuracy of the deformation field estimation. Second, LungRegNet incorporates enhanced lung vessel images as anatomical constraints during the registration process, which improves the alignment precision of the lung vascular regions to some extent. This approach provides an important insight for future research: integrating anatomical information about the lung, such as blood vessels and airways, can significantly improve registration accuracy and robustness, particularly when dealing with complex deformations and local anatomical features.

2 Model Generalization

In this study, MS-VMANet is primarily applied to the registration of lung 4DCT images. However, its architecture exhibits a certain degree of generalizability and, in theory, can be extended to image registration tasks for other anatomical regions, such as the brain, liver, and knee. Since different anatomical regions involve distinct regions of interest (ROIs) during the registration process, the model's performance may be significantly influenced by variations in data characteristics and deformation patterns. Future research could integrate a multi-task learning framework to allow the model to dynamically acquire deformation features from various anatomical regions. Additionally, incorporating adjustable deformation field constraints, such as those informed by anatomical prior knowledge, could improve the model's flexibility and resilience in handling diverse registration tasks.

3 Lack of Multimodal Data and Clinical Validation

This study primarily utilizes 4DCT images of lung cancer patients; however, it does not incorporate other imaging modalities, such as PET or MRI. Integrating multimodal data can provide more comprehensive physiological and anatomical information, thereby improving registration accuracy and enhancing the model′s robustness when handling complex cases. Future research should explore the integration of multimodal datasets, like the VAMPIRE dataset [27], the Learn2Reg dataset [20], and the ChestX-ray8 dataset [39], to further enhance the robustness and generalization of registration methods. Moreover, although the regional ventilation maps obtained in this study are generated based on 4DCT images of lung cancer patients and therefore possess a certain degree of validity, their clinical effectiveness and practical significance have not yet been clinically validated. Therefore, future work should incorporate real ventilation imaging modalities (e. g., SPECT or 68Ga-based PET images) to validate and compare these regional ventilation maps.

6. Conclusion

In this study, we present a multi-scale efficient VMamba attention registration network (MS-VMANet), designed to evaluate ventilation across different lung 4DCT regions. We compared the proposed MS-VMANet approach with other unsupervised learning-based registration approaches and assessed its performance on the publicly available DIRLAB dataset. MS-VMANet achieved significant results in terms of quantification metrics (TRE, DSC, and SSIM), demonstrating its reliability and accuracy. Furthermore, based on the registration results, we quantified the deformation field using the Jacobian determinant to generate accurate functional maps reflecting the ventilation of each lung region. Beyond the evaluation of ventilation, these functional maps can be utilized in a range of clinical contexts, including assisting in radiotherapy planning to avoid regions with high ventilation, monitoring lung function post-treatment, and forecasting potential thoracic complications. This approach offers feasible support for the diagnosis and management of pulmonary disorders.

Acknowledgement

This work was supported by the National Key Research and Development Plan "NQI Integrated Services Common Technologies Research" special project (No. 2021YFF0600100) and the Major Science and Technology Projects of Zhejiang Province (No. 2023C3107).

References

1. Akira, M., Toyokawa, K., Inoue, Y., Arai, T. Quantitative CT in Chronic Obstructive Pulmonary Disease: Inspiratory and Expiratory Assessment. American Journal of Roentgenology, 2009, 192(1), 267-272.https://doi.org/10.2214/AJR.07.3953
2. Arroyo-Hernández, M., Maldonado, F., Lozano-Ruiz, F., Muñoz-Montaño, W., Nuñez-Baez, M., Arrieta, O. Radiation-Induced Lung Injury: Current Evidence. BMC Pulmonary Medicine, 2021, 21, 1-12. https://doi.org/10.1186/s12890-020-01376-4

3. Bajcsy, R., Kovačič, S. Multiresolution Elastic Matching. Computer Vision, Graphics, and Image Processing, 1989, 46(1), 1-21. https://doi.org/10.1016/S0734-189X(89)80014-3

4. Balakrishnan, G., Zhao, A., Sabuncu, M. R., Guttag, J., Dalca, A. V. VoxelMorph: A Learning Framework for Deformable Medical Image Registration. IEEE Transactions on Medical Imaging, 2019, 38(8), 1788-1800. https://doi.org/10.1109/TMI.2019.2897538

5. Castillo, E., Castillo, R., Martinez, J., Shenoy, M., Guerrero, T. Four-Dimensional Deformable Image Registration Using Trajectory Modeling. Physics in Medicine & Biology, 2009, 55(1), 305. https://doi.org/10.1088/0031-9155/55/1/018

6. Chen, J., Frey, E. C., He, Y., Segars, W. P., Li, Y., Du, Y. TransMorph: Transformer for Unsupervised Medical Image Registration. Medical Image Analysis, 2022, 82, 102615. https://doi.org/10.1016/j.media.2022.102615

7. Chen, Z., Zheng, Y., Gee, J. C. TransMatch: A Transformer-Based Multilevel Dual-Stream Feature Matching Network for Unsupervised Deformable Image Registration. IEEE Transactions on Medical Imaging, 2023, 43(1), 15-27.https://doi.org/10.1109/TMI.2023.3288136

8. Clark, K., Vendt, B., Smith, K., Freymann, J., Kirby, J., Koppel, P., Moore, S., Phillips, S., Maffitt, D., Pringle, M., Tarbox, L., Prior, F. The Cancer Imaging Archive (TCIA), Maintaining and Operating a Public Information Repository. Journal of Digital Imaging, 2013, 26(6), 1045-1057.https://doi.org/10.1007/s10278-013-9622-7

9. Ding, K., Bayouth, J. E., Buatti, J. M., Christensen, G. E., Reinhardt, J. M. 4DCT-Based Measurement of Changes in Pulmonary Function Following a Course of Radiation Therapy. Medical Physics, 2010, 37(3), 1261-1272. https://doi.org/10.1118/1.3312210

10. Eijnatten, M. V., Rundo, L., Batenburg, K. J., Lucka, F., Beddowes, E., Caldas, C., Gallagher, F. A., Sala, E., Schönlieb, C. B., Woitek, R. 3D Deformable Registration of Longitudinal Abdominopelvic CT Images Using Unsupervised Deep Learning. Computer Methods and Programs in Biomedicine, 2021, 208, 106261. https://doi.org/10.1016/j.cmpb.2021.106261

11. Fakhfakh, M., Bouaziz, B., Gargouri, F., Chaari, L. ProgNet, COVID-19 Prognosis Using Recurrent and Convolutional Neural Networks. The Open Medical Imaging Journal, 2020. https://doi.org/10.1101/2020.05.06.20092874

12. Fang, Q., Gu, X., Yan, J., Zhao, J., Li, Q. A FCN-Based Unsupervised Learning Model for Deformable Chest CT Image Registration. 2019 IEEE Nuclear Science Symposium and Medical Imaging Conference (NSS/MIC), 2019, 1-4. https://doi.org/10.1109/NSS/MIC42101.2019.9059976

13. Foote, M. D., Zimmerman, B. E., Sawant, A., Joshi, S. C. Real-Time 2D-3D Deformable Registration with Deep Learning and Application to Lung Radiotherapy Targeting. Information Processing in Medical Imaging: 26th International Conference (IPMI), 2019, 11492, 265-276. https://doi.org/10.1007/978-3-030-20351-1_20

14. Fu, Y., Lei, Y., Wang, T., Higgins, K., Bradley, J. D., Curran, W. J., Liu, T., Yang, X. LungRegNet, An Unsupervised Deformable Image Registration Method for 4DCT Lung. Medical Physics, 2020, 47(4), 1763-1774. https://doi.org/10.1002/mp.14065

15. Guerrero, T., Sanders, K., Castillo, E., Zhang, Y., Bidaut, L., Pan, T., Komaki, R. Dynamic Ventilation Imaging from Four-Dimensional Computed Tomography. Physics in Medicine & Biology, 2006, 51(4), 777-791. https://doi.org/10.1088/0031-9155/51/4/002

16. Hansen, L., Heinrich, M. P. GraphRegNet: Deep Graph Regularisation Networks on Sparse Keypoints for Dense Registration of 3D Lung CTs. IEEE Transactions on Medical Imaging, 2021, 40(9), 2246-2257. https://doi.org/10.1109/TMI.2021.3073986

17. Hassani, A., Shi, H. Dilated Neighborhood Attention Transformer. Computer Vision and Pattern Recognition, 2022. https://doi.org/10.1109/CVPR52729.2023.00599

18. He, X., Guo, J., Zhang, X., Bi, H., Laine, A. Recursive Refinement Network for Deformable Lung Registration Between Exhale and Inhale CT Scans. arXiv preprint, 2021. https://doi.org/10.48550/arXiv.2106.07608

19. Heinrich, M. P., Hansen, L. VoxelMorph++ Going Beyond the Cranial Vault with Keypoint Supervision and Multi-Channel Instance Optimization. International Workshop on Biomedical Image Registration, 2022, 13386, 85-95. https://doi.org/10.1002/mp.16548

20. Hering, A., Hansen, L., Mok, T. C. W., Chung, A. C. S., Siebert, H., Häger, S. Learn2Reg, Comprehensive Multi-Task Medical Image Registration Challenge, Dataset and Evaluation in the Era of Deep Learning. IEEE Transactions on Medical Imaging, 2022, 42(3), 697-712. https://doi.org/10.1109/TMI.2022.3213983

21. Hering, A., van Ginneken, B., Heldmann, S. MLViRNet: Multilevel Variational Image Registration Network. Medical Image Computing and Computer-Assisted Intervention (MICCAI), 2019, 11769, 257-265. https://doi.org/10.1007/978-3-030-32226-7_29

22. Iandola, F., Moskewicz, M., Karayev, S., Girshick, R., Darrell, T., Keutzer, K. Densenet: Implementing Efficient Convnet Descriptor Pyramids. arXiv preprint, 2014. https://doi.org/10.48550/arXiv.1404.1869

23. Iqbal, M. Z., Razzak, I., Qayyum, A., Nguyen, T. T., Tanveer, M., Sowmya, A. Hybrid Unsupervised Paradigm Based Deformable Image Fusion for 4D CT Lung Image Modality. Information Fusion, 2024, 102, 102061. https://doi.org/10.1016/j.inffus.2023.102061

24. Jaderberg, M., Simonyan, K., Zisserman, A. Spatial Transformer Networks. Advances in Neural Information Processing Systems (NeurIPS), 2015, 28.

25. Kang, M., Hu, X., Huang, W., Scott, M. R., Reyes, M. Dual-Stream Pyramid Registration Network. Medical Image Analysis, 2022, 78, 102379. https://doi.org/10.1016/j.media.2022.102379

26. Kipritidis, J., Siva, S., Hofman, M. S., Callahan, J., Hicks, R. J., Keall, P. J. Validating and Improving CT Ventilation Imaging by Correlating with Ventilation 4D-PET/CT Using 68Ga-Labeled Nanoparticles. Medical Physics, 2014, 41(1), 011910. https://doi.org/10.1118/1.4856055

27. Kipritidis, J., Tahir, B. A., Cazoulat, G., Hofman, M. S., Siva, S., Callahan, J., Hardcastle, N., Yamamoto, T., Christensen, G. E., Reinhardt, J. M., Kadoya, N., Patton, T. J., Gerard, S. E., Duarte, I., Archibald-Heeren, B., Byrne, M., Sims, R., Ramsay, S., Booth, J. T., Eslick, E., Hegi-Johnson, F., Woodruff, H. C., Ireland, R. H., Wild, J. M., Cai, J., Bayouth, J. E., Brock, K., Keall, P. J. The VAMPIRE Challenge, A Multi-Institutional Validation Study of CT Ventilation Imaging. Medical Physics, 2019, 46(3), 1198-1217. https://doi.org/10.1002/mp.13346

28. Kratzer, T. B., Bandi, P., Freedman, N. D., Smith, R. A., Travis, W. D., Jemal, A., Siegel, R. L. Lung Cancer Statistics, 2023. Cancer, 2024, 130(8), 1330-1348. https://doi.org/10.1002/cncr.35128

29. Kybic, J., Unser, M. Fast Parametric Elastic Image Registration. IEEE Transactions on Image Processing, 2003, 12(11), 1427-1442. https://doi.org/10.1109/TIP.2003.813139

30. Li, M., Castillo, E., Zheng, X. L., Luo, H. Y., Castillo, R., Wu, Y., Guerrero, T. Modeling Lung Deformation: A Combined Deformable Image Registration Method with Spatially Varying Young's Modulus Estimates. Medical Physics, 2013, 40(8), 081902. https://doi.org/10.1118/1.4812419

31. Lu, J., Jin, R., Song, E., Ma, G., Wang, M. Lung-CRNet: A Convolutional Recurrent Neural Network for Lung 4DCT Image Registration. Medical Physics, 2021, 48(12), 7900-7912. https://doi.org/10.1002/mp.15324

32. Mok, T. C. W., Chung, A. C. S. Large Deformation Diffeomorphic Image Registration with Laplacian Pyramid Networks. International Conference on Medical Image Computing and Computer-Assisted Intervention (MICCAI), 2020, 12263, 211-221. https://doi.org/10.1007/978-3-030-59716-0_21

33. Monji-Azad, S., Kinz, M., Kothari, S., Khanna, R., Mihan, A. C., Maennel, D., Scherl, C., Hesser, J. DefTransNet: A Transformer-Based Method for Non-Rigid Point Cloud Registration in the Simulation of Soft Tissue Deformation. arXiv preprint, 2025. https://doi.org/10.1088/1361-6501/ade613

34. Paszke, A., Gross, S., Massa, F., Lerer, A., Bradbury, J., Chanan, G., Killeen, T., Lin, Z., Gimelshein, N., Antiga, L., Desmaison, A., Kopf, A., Yang, E., DeVito, Z., Raison, M., Tejani, A., Chilamkurthy, S., Steiner, B., Fang, L., Bai, J., Chintala, S. PyTorch, An Imperative Style, High-Performance Deep Learning Library. Advances in Neural Information Processing Systems, 2019, 32. https://proceedings.neurips.cc/

35. Peter, L., Alexander, D. C., Magnain, C., Iglesias, J. E. Uncertainty-Aware Annotation Protocol to Evaluate Deformable Registration Algorithms. IEEE Transactions on Medical Imaging, 2021, 40(8), 2053-2065. https://doi.org/10.1109/TMI.2021.3070842

36. Reinhardt, J. M., Christensen, G. E., Hoffman, E. A., Ding, K., Cao, K. Registration-Derived Estimates of Local Lung Expansion as Surrogates for Regional Ventilation. Biennial International Conference on Information Processing in Medical Imaging, 2007, 4580, 763-774. https://doi.org/10.1007/978-3-540-73273-0_63

37. Reinhardt, J. M., Ding, K., Cao, K., Christensen, G. E., Hoffman, E. A., Bodas, S. V. Registration-Based Estimates of Local Lung Tissue Expansion Compared to Xenon CT Measures of Specific Ventilation. Medical Image Analysis, 2008, 12(6), 752-763. https://doi.org/10.1016/j.media.2008.03.007

38. Shaker, A. M., Maaz, M., Rasheed, H., Khan, S., Yang, M. H., Khan, F. S. UNETR++, Delving into Efficient and Accurate 3D Medical Image Segmentation. IEEE Transactions on Medical Imaging, 2024, 43(9), 3377-3390. https://doi.org/10.1109/TMI.2024.3398728

39. Shiraishi, J., Katsuragawa, S., Ikezoe, J., Matsumoto, T., Kobayashi, T., Komatsu, K., Matsui, M., Fujita, H., Kodera, Y., Doi, K. Development of a Digital Image Database for Chest Radiographs with and without a Lung Nodule, Receiver Operating Characteristic Analysis of Radiologists' Detection of Pulmonary Nodules. American Journal of Roentgenology, 2000, 174(1), 71-74. https://doi.org/10.2214/ajr.174.1.1740071

40. Sim, J. K., Moon, S. J., Choi, J., Oh, J. Y., Lee, Y. S., Min, K. H., Hur, G. Y., Lee, S. Y., Shim, J. J. Mechanical Ventilation in Patients with Idiopathic Pulmonary Fibrosis in Korea, A Nationwide Cohort Study. The Korean Journal of Internal Medicine, 2024, 39(2), 295. https://doi.org/10.3904/kjim.2023.273

41. Teng, X., Chen, Y., Zhang, Y., Ren, L. Respiratory Deformation Registration in 4DCT/Cone Beam CT Using Deep Learning. Quantitative Imaging in Medicine and Surgery, 2021, 11(2), 737. https://doi.org/10.21037/qims-19-1058

42. Thirion, J. P. Image Matching as a Diffusion Process: An Analogy with Maxwell's Demons. Medical Image Analysis, 1998, 2(3), 243-260. https://doi.org/10.1016/S1361-8415(98)80022-4

43. Vandemeulebroucke, J., Rit, S., Kybic, J., Clarysse, P., Sarrut, D. Spatiotemporal Motion Estimation for Respiratory-Correlated Imaging of the Lungs. Medical Physics, 2011, 38(1), 166-178. https://doi.org/10.1118/1.3523619

44. Vinogradskiy, Y., Castillo, R., Castillo, E., Schubert, L., Jones, B. L., Faught, A., Gaspar, L. E., Kwak, J., Bowles, D. W., Waxweiler, T., Dougherty, J. M., Gao, D., Stevens, C., Miften, M., Kavanagh, B., Grills, I., Rusthoven, C. G., Guerrero, T. Results of a Multi-Institutional Phase 2 Clinical Trial for 4DCT-Ventilation Functional Avoidance Thoracic Radiation Therapy. International Journal of Radiation Oncology Biology Physics, 2022, 112(4), 986-995. https://doi.org/10.1016/j.ijrobp.2021.10.147

45. Wang, Y., Solomon, J. M. PRNet: Self-Supervised Learning for Partial-to-Partial Registration. Advances in Neural Information Processing Systems (NeurIPS), 2019, 32, 8761-9553.

46. Wong, Y. M., Yeap, P. L., Ong, A. L. K., Tuan, J. K. L., Lew, W. S., Lee, J. C. L., Tan, H. Q. Machine Learning Prediction of Dice Similarity Coefficient for Validation of Deformable Image Registration. Intelligence-Based Medicine, 2024, 10, 100163. https://doi.org/10.1016/j.ibmed.2024.100163

47. Zhao, S., Dong, Y., Chang, E. I., Xu, Y. Recursive Cascaded Networks for Unsupervised Medical Image Registration. Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV), 2019, 10600-10610. https://doi.org/10.1109/ICCV.2019.01070

48. Zhu, L., Liao, B., Zhang, Q., Wang, X., Liu, W., Wang, X. Vision Mamba, Efficient Visual Representation Learning with Bidirectional State Space Model. Advances in Neural Information Processing Systems, 2024, 37, 103031-103063. https://arxiv.org/html/2401.09417v1