

Comprehensive Quality Management Solutions



SunSCAN™ 3D Water Scanning System



SunCHECK® Platform

Demo these smarter Quality Management solutions at the ASTRO Annual Meeting. Join Sun Nuclear in Booth #1835 for in-booth talks, product demos, and more.

Learn more:
sunnuclear.com/astro



SRS MR Distortion Phantom



StereoPHAN™, SRS MapCHECK®, MultiMet-WL Cube

sunnuclear.com

The SunSCAN 3D system is not available for sale in all markets. CE mark pending.



SUN NUCLEAR
A MIRION MEDICAL COMPANY

RESEARCH ARTICLE

MEDICAL PHYSICS

Deep-learning based fast and accurate 3D CT deformable image registration in lung cancer

Yuzhen Ding¹ | Hongying Feng¹ | Yunze Yang¹ | Jason Holmes¹ |
 Zhengliang Liu² | David Liu³ | William W. Wong¹ | Nathan Y. Yu¹ |
 Terence T. Sio¹ | Steven E. Schild¹ | Baoxin Li⁴ | Wei Liu¹

¹Department of Radiation Oncology, Mayo Clinic, Phoenix, Arizona, USA

²Department of Computer Science, University of Georgia, Athens, Georgia, USA

³Athens Academy, Athens, Georgia, USA

⁴School of Computing and Augmented Intelligence, Arizona State University, Tempe, Arizona, USA

Correspondence

Wei Liu, Department of Radiation Oncology, Mayo Clinic Arizona, 5777 E. Mayo Boulevard, Phoenix, AZ 85054, USA.
 Email: Liu.Wei@mayo.edu

Funding information

National Cancer Institute (NCI), Grant/Award Number: K25CA168984

Abstract

Background: Deformable Image Registration (DIR) is an essential technique required in many applications of radiation oncology. However, conventional DIR approaches typically take several minutes to register one pair of 3D CT images and the resulting deformable vector fields (DVF) are only specific to the pair of images used, making it less appealing for clinical application.

Purpose: A deep-learning-based DIR method using CT images is proposed for lung cancer patients to address the common drawbacks of the conventional DIR approaches and in turn can accelerate the speed of related applications, such as contour propagation, dose deformation, adaptive radiotherapy (ART), etc.

Methods: A deep neural network based on VoxelMorph was developed to generate DVFs using CT images collected from 114 lung cancer patients. Two models were trained with the weighted mean absolute error (wMAE) loss and structural similarity index matrix (SSIM) loss (optional) (i.e., the MAE model and the M+S model). In total, 192 pairs of initial CT (*iCT*) and verification CT (*vCT*) were included as a training dataset and the other independent 10 pairs of CTs were included as a testing dataset. The *vCT*s usually were taken 2 weeks after the *iCT*s. The synthetic CTs (*sCT*s) were generated by warping the *vCT*s according to the DVFs generated by the pre-trained model. The image quality of the synthetic CTs was evaluated by measuring the similarity between the *iCT*s and the *sCT*s generated by the proposed methods and the conventional DIR approaches, respectively. Per-voxel absolute CT-number-difference volume histogram (CDVH) and MAE were used as the evaluation metrics. The time to generate the *sCT*s was also recorded and compared quantitatively. Contours were propagated using the derived DVFs and evaluated with SSIM. Forward dose calculations were done on the *sCT*s and the corresponding *iCT*s. Dose volume histograms (DVHs) were generated based on dose distributions on both *iCT*s and *sCT*s generated by two models, respectively. The clinically relevant DVH indices were derived for comparison. The resulted dose distributions were also compared using 3D Gamma analysis with thresholds of 3 mm/3%/10% and 2 mm/2%/10%, respectively.

Results: The two models (wMAE and M+S) achieved a speed of 263.7 ± 163 / 265.8 ± 190 ms and a MAE of 13.15 ± 3.8 / 17.52 ± 5.8 HU for the testing dataset, respectively. The average SSIM scores of 0.987 ± 0.006 and 0.988 ± 0.004 were achieved by the two proposed models, respectively. For both models, CDVH of a typical patient showed that less than 5% of the voxels had a per-voxel absolute CT-number-difference larger than 55 HU. The dose distribution calculated based

on a typical sCT showed differences of $\leq 2\text{cGy[RBE]}$ for clinical target volume (CTV) D_{95} and D_5 , within $\pm 0.06\%$ for total lung V_5 , $\leq 1.5\text{cGy[RBE]}$ for heart and esophagus D_{mean} , and $\leq 6\text{cGy[RBE]}$ for cord D_{max} compared to the dose distribution calculated based on the iCT. The good average 3D Gamma passing rates ($> 96\%$ for 3 mm/3%/10% and $> 94\%$ for 2 mm/2%/10%, respectively) were also observed.

Conclusion: A deep neural network-based DIR approach was proposed and has been shown to be reasonably accurate and efficient to register the initial CTs and verification CTs in lung cancer.

KEYWORDS

3D lung CT images, deep neural networks, deformable image registration

1 | INTRODUCTION

Image registration aims to find the spatial relationship between two or multiple sets of images and is usually formalized as the optimization of a function balancing the similarity between images (either in intensity, topology, or both).¹ Compared to rigid image registration (RIR), deformable image registration (DIR) attempts to find the voxel-specific spatial relationship between two or multiple sets of images. Therefore, DIR has far more flexibilities than RIR and can be used in more complicated clinical scenarios such as images with large anatomical structure changes. DIR has been extensively used in radiation therapy¹ such as automatic segmentation,^{2,3} mathematical modeling,^{4–7} functional imaging,^{8–10} and dose deformation.^{11–16}

Over the years, many conventional DIR approaches have been developed and adopted clinically. The conventional DIR approaches can be broadly categorized into two categories: parametric^{6,7,17} and non-parametric models.^{18–21} The parametric model generates DVFs as a linear combination of its basic functions. The B-spline model^{22–25} is an example of such parametric models and it can handle the local change of a voxel by linear regression from nearby voxels within a certain distance. This property significantly reduces the computation time and memory required. For example, Shekhar et al.²⁶ proposed a DIR framework for auto-segmentation. The framework consists of a B-spline-based transformation model, mean squared difference-based image similarity measure, and a downhill simplex algorithm as the optimization scheme. It achieved fewer than 120 and 135HU mean squared difference for lung and abdomen patients, respectively. Yet, the results can only be used for CTs with either breath-holding or respiratory gating, which limit its wide applications in clinics. In contrast, non-parametric models such as demons-based^{18–21} methods calculate transformation vectors of all voxels, thus achieving more accurate DVFs, but requiring more computation time and memory than the parametric models. For example, Reed et al.²⁷ achieved an average of 1.3 mm mean displacement in auto-segmentation for

10 patients using an accelerated “demons” algorithm,²⁸ which adds a HU number gradient similarity term and a transformation error term into the demons’ energy function, and uses the limited Broyden-Fletcher-Goldfarb-Shanno (L-BFGS) algorithm²⁹ to automatically determine the iteration number, thus accelerating the algorithm. However, it also requires the patients to have a similar body mass index (BMI),²⁹ which also potentially limits its application clinically.

Modern radiation therapy is increasingly sophisticated with more beam delivery techniques such as intensity modulation and/or volumetric modulation, including intensity modulated X-ray-based radiation therapy (IMRT),^{30–33} volumetric modulated arc therapy (VMAT),³⁴ and intensity-modulated proton therapy (IMPT).^{35–43} IMPT enjoys distinct advantages in terms of high conformality of target coverage and superior organs-at-risk (OARs) protection owing to its high flexibility at the beamlet level in treatment planning and dose delivery.^{35–38} However, it is also extremely sensitive to proton beam range, patient setup uncertainties, intra- and inter- fractional anatomical changes.^{38,44–78} The concept of adaptive radiotherapy (ART)^{14,79,80} has been introduced to account for anatomical changes during treatment course. For ART, patients under treatment require periodic verification imaging during treatment course to obtain information about their internal anatomical changes. However, the potential gain of ART is at the cost of increasing clinical workload, such as CT deformation, contour propagation, and dose deformation. Those clinical tasks all depend on the availability and quality of DIR. Unfortunately, it typically takes minutes for the conventional DIR approaches to register one pair of 3D CTs and the resulted deformable vector fields (DVFs) are not generalized to other CT images, even when they are similar or from the same patient, hence greatly limiting its further applications in ART, which is very time sensitive. Moreover, the frequency of re-planning is significantly higher in IMPT than IMRT/VMAT (e.g., for head and neck cancer, 20–25% for IMRT/VMAT and 45–50% for IMPT). This makes the same tasks even more labor intensive in proton clinics. Therefore, the

undesired patient breaks allowing for tumor cell repopulation might take place at busy clinics due to insufficient resources.

Recently, several deep learning-based methods have been developed to speed up DIR in medical image analysis.^{81–83} Yang et al.⁸¹ proposed a two-steps deep learning framework for predicting the momentum parameterization for the large deformation diffeomorphic metric mapping (LDDMM) model. The proposed deep learning framework consists of two auto-encoder networks with the same architecture, in which the first auto-encoder is used to estimate the initial patch-wise momentum and the second one further tunes the initial patch-wise momentum. Although the proposed method is much faster compared to the conventional DIR approaches, the computational complexity is higher than a typical single-step deep learning network. Besides, since it has two cascade networks, the symmetrical error may accumulate as the layers go deeper. Balakrishnan et al.⁸² proposed a UNet-like model termed as VoxelMorph to learn the DVFs from pairs of magnetic resonance images (MRIs) (i.e., moving images and fixed images), then the generated DVFs and moving images go through a non-learnable spatial transformation to form the final generated warped images that resemble the fixed images. The VoxelMorph can achieve comparable performance as the state-of-the-art conventional DIR methods, whereas it is orders of magnitude faster. Thus, it has been widely used in medical image analysis. Most of these methods have been proposed for MRIs, which typically have high-resolution and rich anatomical information, whereas in radiation therapy the commonly used image modality is CT with a relatively low resolution. Vos et al.⁸³ proposed a deep learning image registration (DLIR) framework for unsupervised affine and deformable image registration. It uses convolutional layers to predict the B-spline control points in each of the three directions, then the DVFs are generated from the estimated control points by B-spline interpolation, which is implemented by transpose convolutions. Although the DLIR can be applied to both MRIs and CTs, when it is used for CT images, it requires a large training dataset to train the model. Moreover, the trained model can only be used for 4D CTs with only intra-fractional anatomical changes considered, which limits its clinical use. Another deep learning-based DIR approach was proposed by Zhao et al.⁸⁴ by cascading multiple Volume Tweening Network (VTN) networks to recursively generate coarse-to-fine DVFs. Typically, the more the cascades are, the more accurate the generated DVFs are. However, the number of cascades is bounded by the GPU memory, and a large amount of data is required to train such a large-scale network, which is challenging for tasks involving medical images.

To address the aforementioned challenges for the deep learning-based DIR approaches to be used in CTs (e.g., dependence of large training dataset, limitation of

4D CTs, and requirements of high resolution, which is not available in CTs), we proposed several additional loss terms in the objective function of VoxelMorph as well as a random masking strategy to greatly improved the quality of the synthetic CT images (a similar idea has been adopted by He et al.⁸⁵ to significantly accelerate training speed as well as improve the classification accuracy), yielding an efficient, accurate, and generalizable deep-learning based DIR method for CTs.

The contributions can be summarized as follows:

1. We proposed a novel VoxelMorph-based framework for inter-fractional lung DIR. Different from conventional DIR approaches that are only specific to the images used, our framework can be generalized to any images of any independent patients once the model is trained. Thus, it is more practical and versatile.
2. A new random masking strategy was proposed to significantly reduce artifacts in the deformed images due to intrinsic low resolution of the CT images compared with MRIs. In addition, we investigated the functionalities of different loss terms used in the model training and used weighted mean absolute error (wMAE) and structural similarity index matrix (SSIM) loss (optional) to bridge the gap between CT images and MRIs, the latter has been well studied in deep learning-based DIR. Thus, the image quality of the deformed CTs is further improved.
3. Dedicated pre- and post-process methods are proposed to standardize all the CT images used in this work. Then, as a demonstration, we constructed a novel diversified inter-fractional lung CT dataset consisting of approximately 200 pairs of such standardized CT images collected from patients treated by both proton therapy and photon therapy. Such a dataset can be used to evaluate the performance of not only the DIR approaches, but also other related tasks. In the meantime, the proposed pre- and post-process methods can be applied to other medical images (e.g., head and neck CT images, MRI, etc.).
4. Our methods achieved the state-of-the-art performance in terms of time efficiency, high reconstructed image quality, indistinguishable dose distribution difference calculated between the ground-truth and deformed CTs, and good Gamma passing rates.

2 | METHODS

To address the drawbacks of the conventional DIR approaches, such as low accuracy and large time consumption, we propose to train a deep-learning-based model for the deformable vector fields (DVFs) with VoxelMorph, which is a general-purpose library for deep-learning-based tools for registration and deformations and includes two additional loss terms that focus

on voxel-level similarity and structure-level similarity, respectively. We also introduce a new training strategy that can alleviate the artifacts in low resolution images (i.e., CT images) thus achieving accurately warped CT images. In Section 2.1, we describe the data collection and data preprocessing. Section 2.2 introduces the overall structure of the proposed method and in Section 2.3, the training process and validation process of the proposed method are elaborated. Statistical analysis is included in Section 2.4.

2.1 | Data pre-processing

The initial CT (iCT_0) and verification CT (vCT_0) of 114 lung cancer patients treated at our institution were retrospectively selected, among which the CT images from 104 patients were used for training and 10 were used for testing. Each patient had one initial CT and 1–4 verification CTs, forming a training dataset of 192 pairs of CT images and a testing dataset of 10 pairs of CT images. In the training dataset, 101 pairs were collected from 67 patients treated with photon therapy whereas the other 91 pairs were collected from 37 patients treated with proton therapy. Among the 10 testing patients, seven patients were treated with photon therapy, while the other three were treated with proton therapy.

As the collected CT images were captured at various times and by various CT simulators, the CT images may be different due to anatomical changes and various configurations of the different CT simulators. To make sure that the dataset was consistent, data preprocessing was conducted as follows. First, we used the iterative metal artifact reduction (iMAR) algorithm,⁸⁶ which is integrated in the commercial software for the CT simulator, to remove artifacts caused by metal implants. Then, rigid registration and center-cropping were applied to all the CT images using the following technique: we first randomly picked one CT image, where the regions of interest (ROIs) were roughly located in the center of the 3D CTs. We regarded this CT image set as the reference CT (rCT). Then, we registered (rigid) all other CTs to rCT using the Insight Toolkit (ITK)⁸⁷ such that all CTs had the same resolution of $2 \times 1.26 \times 1.26 \text{ mm}^3$, the same dimension size, etc. Next, we center-cropped all CTs to a dimension size of $136 \times 384 \times 384$ to exclude the non-informative areas from this study as well as to alleviate the memory burden in training. We manually selected a fixed center-cropping region instead of using the BODY contour since the BODY contour varied from patient to patient, and in some proton plans the BODY contour contained the digital couch. Last, we normalized all the CT numbers to values approximately around 0 to 1 by using a uniform shift of 1000 and a fixed denominator of 3000. The preprocessed initial CT (iCT) and verification CT (vCT) were then used for the model training and validation. The workflow of both the data pre-processing and data post-processing steps is shown in Figure 1.

2.2 | Overview of the proposed framework

An overview of the proposed framework is illustrated in Figure 2 and the detailed model architecture is shown in Figure 3. The input for the model is a pair of CT images, iCT and vCT , that are taken at different time points (usually several weeks apart). The backbone of the model is VoxelMorph, which is a UNet-like deep neural network architecture and the output from the network is the DVFs. Finally, the vCT undergoes a spatial transformation based on the derived DVFs to form the final output—the sCT (Figure 2).

Since the 3D lung CT images have different resolutions as well as different dimension sizes as the MRIs used in the original VoxelMorph model, the kernel size, stride, and other parameters are changed accordingly to make sure that the CT images and the network are compatible. To be more specific (Figure 3), 3D convolutional layers are used with a kernel size of 3 and a stride of 1 in both the encoder and decoder. A LeakyReLU⁸⁸ layer with a parameter of 0.2 was applied right after each convolutional layer. The convolutional layers together with down-sampling across different layers allow us to capture the hierarchical features, which are derived from the input CT image pairs. Similarly, the decoder learns the DVFs from both the hierarchical features extracted by each layer in the encoder and the previous layer in the decoder. To deal with the odd number of the feature maps in the deepest layer of the encoder and decoder, we randomly duplicated one of the feature maps and concatenated it with the original feature maps (the second layer in the decoder), thus the number of the feature maps is consistent in both the encoder and decoder. Finally, the output from the network, that is, the DVFs, were applied to the vCT image to generate the sCT image through spatial transform, in which the voxel location is first calculated then followed by linear interpolation. The sCT quality is evaluated with iCT as the ground-truth.

Both vCT and DVF, which is generated by the model (step 2), will go through the spatial transformation (step 3) to obtain the final output sCT (step 4). The training and inference path have been indicated by bolded numbers. Note that the random mask (red rectangle block in both iCT and vCT) is only applied in the training stage. More details about random mask will be introduced in Section 2.3.

2.3 | Training and validation protocols of the proposed framework

As shown in Figures 2 and 3, the base architecture is the VoxelMorph, which is a UNet-like structure that was proposed for DIR of the MRI images in head and neck. Although the vanilla VoxelMorph works well for DIR of the MRI images in head and neck, its

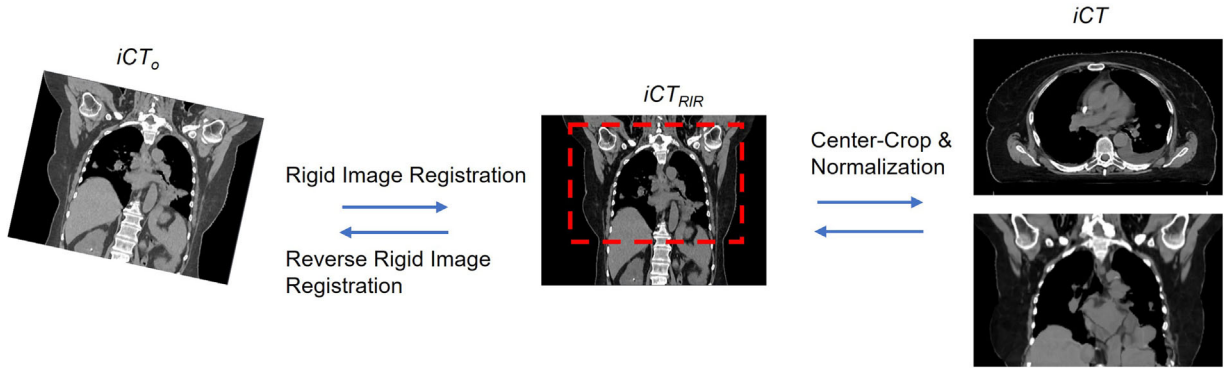


FIGURE 1 Illustration of the data pre-processing and post-processing steps. iCT_0 represented the raw data, iCT_{RIR} represented the CT images after RIR was applied to iCT_0 and iCT were the images that we used in the model training and validation.

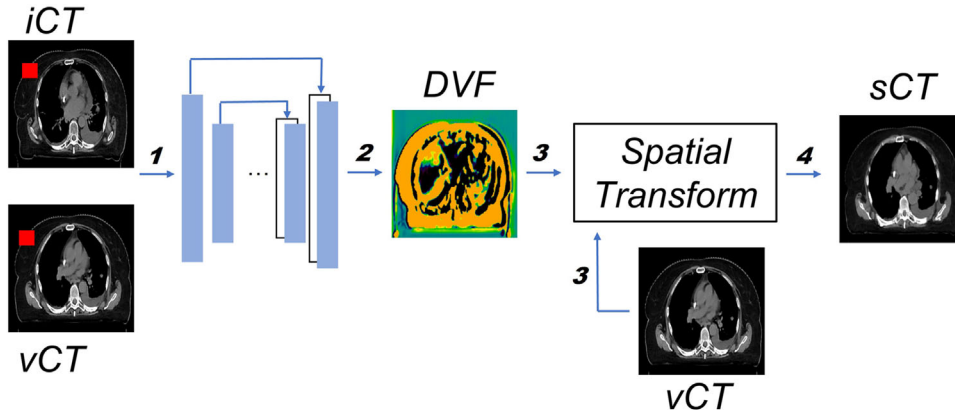


FIGURE 2 Overview of the proposed workflow. The inputs consist of both iCT and vCT (step 1). Both vCT and DVF , which is generated by the model (step 2), will go through the spatial transformation (step 3) to obtain the final output sCT (step 4). The training and inference path have been indicated by bolded numbers. Note that the random mask (red rectangle block in both iCT and vCT) is only applied in the training stage. More details about random mask will be introduced in Section 2.3.

performance greatly degenerated when directly applied to lung CT images. A few reasons contribute to such a degeneration: 1) The MRI images typically have a much higher resolution than the CT images and thus the former will let the model capture more voxel-wise details; 2) The number of the MRI images used in the previous study are much larger than the number of the CT images used in this study to which the model can easily overfit; 3) The lung disease site has larger variation across different patients and within one patient due to inter/intra-fractional anatomy changes than the head and neck disease site. Therefore, to address the above-mentioned challenges, we proposed a new training strategy and a new loss function.

In the training stage, we proposed to use the weighted per-voxel HU number mean absolute error (MAE) ($wMAE$) loss to measure the voxel-wise similarity between the ground-truth iCT s and synthetic sCT s. The definition of the $wMAE$ is:

$$L_{wMAE}(iCT, sCT) = w_{p_{HU}} \sum_{p \in \Omega} |iCT(p) - sCT(p)| \quad (1)$$

where $w_{p_{HU}}$ represents the HU number of the voxel $p \in \Omega$. Different from a plain MAE loss, the weight of the similarity loss of each voxel is proportional to the corresponding HU number of the voxel. Hence, the voxels in structures with higher HU number, for example, bone, were assigned with larger weights than other voxels. This, together with other loss terms help diminish the appearance of high HU number artifacts.

Following VoxelMorph, we also applied the smooth loss term to the generated DVFs, making it physically realistic. The smooth loss term is defined in Equation (2) as follows:

$$L_{smooth}(DVF) = \sum_{p \in \Omega} \|\nabla g(p)\|^2 \quad (2)$$

where $\nabla g(p) = (\frac{\partial g(p)}{\partial x}, \frac{\partial g(p)}{\partial y}, \frac{\partial g(p)}{\partial z})$ is the spatial gradients ∇g of the voxel p . To simplify the computation, we used $\frac{\partial g(p)}{\partial x} \approx g(p_x + 1, p_y, p_z) - g(p_x, p_y, p_z)$, $\frac{\partial g(p)}{\partial y} \approx g(p_x, p_y + 1, p_z) - g(p_x, p_y, p_z)$ and $\frac{\partial g(p)}{\partial z} \approx g(p_x, p_y, p_z + 1) - g(p_x, p_y, p_z)$.

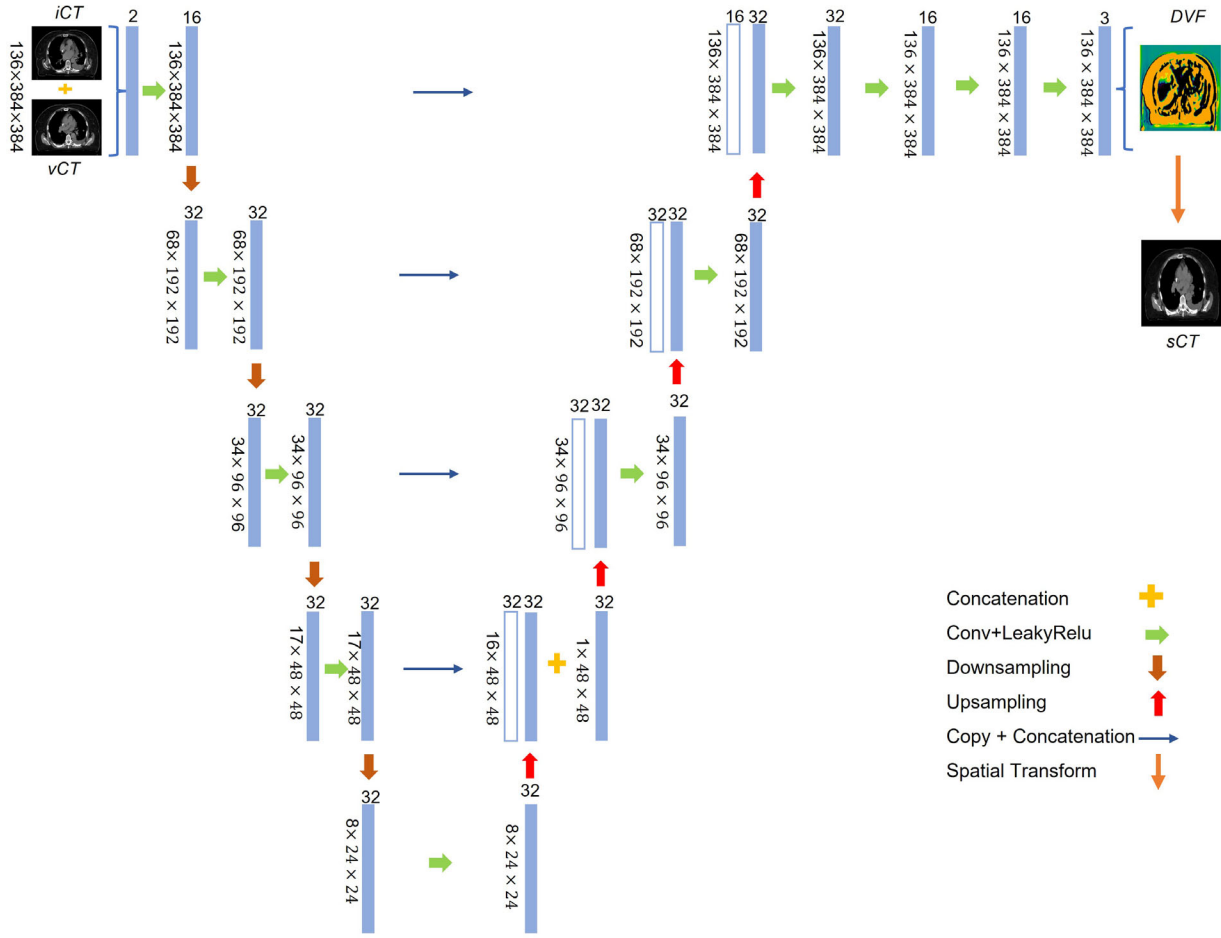


FIGURE 3 Details of the neural network architecture. Every block represents one layer, the value on the left is the input image size, whereas the value on the top indicates the number of feature maps.

$g(p_x, p_y, p_z + 1) - g(p_x, p_y, p_z)$ to approximate the spatial gradients.

Then, we combined all loss terms to obtain the objective function as follows:

$$L = L_{\text{wMAE}} + \alpha L_{\text{smooth}} \quad (3)$$

where α is the weight for the smooth loss term. The model trained with Equation (3) as the objective function is referred to as the wMAE model. In training, α was set to 0.01.

Since a lung CT image typically contains multiple structures with distinct HU numbers, it is challenging to recover all structures simultaneously. Thus, we further extended wMAE model by applying a structure loss term to each of the clinical target volume (CTV) and five organs at risk (OARs), namely, esophagus, heart, left lung, right lung, and cord. To be specific, the contours of both CTV and OARs were converted to bitmaps with 1 indicating the voxels within the ROIs and with 0 indicating the voxels outside the ROIs. Then, the bitmaps of sCT structures were generated by wrapping the bitmaps of vCT structures based on the generated DVF. Last, we used structural similarity index matrix (SSIM)⁸⁹ as

the structure loss to compare the similarity between the bitmaps of sCT and iCT structures. SSIM was applied since it considers not only the similarity between the corresponding structures but also the illumination and contrast of the structures. On the contrary, the commonly used dice similarity coefficients (DSCs)^{22,90,91} only considers the volume overlap of the structures. The SSIM is more appropriate for the lung CT images since the lung CT images often have multiple structures with a large range of the HU numbers that potentially leads to diverse illuminations and contrasts. The definition of SSIM is defined in Equation (4) as follows:

$$L_{\text{SSIM}}(s_{\text{iCT}}, s_{\text{sCT}}) = \frac{1}{K} \sum_{j=1}^K L_{\text{SSIM}}(s_{\text{iCT}}^j, s_{\text{sCT}}^j) \quad (4)$$

$$L_{\text{SSIM}}(x, y) = \frac{(2\mu_x\mu_y + C_1)(2\sigma_{xy} + C_2)}{(\mu_x^2 + \mu_y^2 + C_1)(\sigma_x^2 + \sigma_y^2 + C_2)}$$

where s_{iCT} represents the structures in iCT and s_{sCT} represents the structures in sCT, respectively. K is the number of structures we considered in this study. μ_j and σ_j represents the mean and standard deviation of the

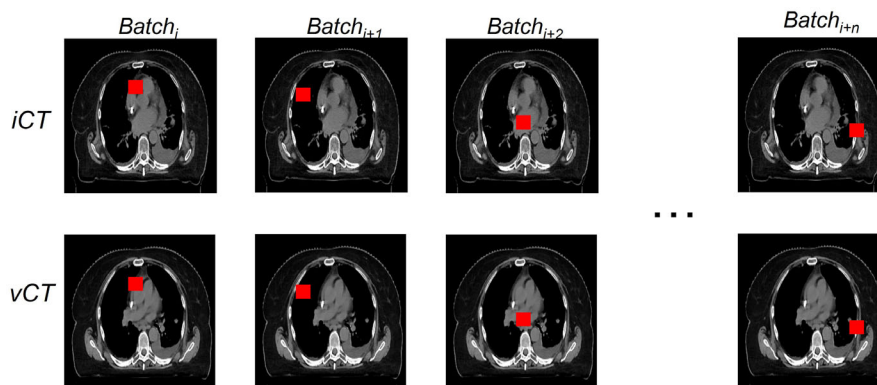


FIGURE 4 Illustration of the random mask strategy used in the model training. In each batch, the value of the voxels enclosed by the random mask will be set to 0.

voxels in the structure j , respectively. C_1 and C_2 are two constants that ensure stability when the denominator becomes 0. A SSIM value of 1 indicates the best agreement and a value of 0 indicates the worst agreement. Finally, we combined all loss terms to obtain the objective function as follows:

$$L = L_{\text{wMAE}} + \alpha L_{\text{smooth}} + \beta L_{\text{SSIM}} \quad (5)$$

where α and β are the weights for different loss. The model trained with Equation (5) as the objective function is referred as the M+S model. In training α was set to 0.01 and β was set to 0.1.

Considering the limited but diverse lung CT images used in this study and the small dimension size of each CT image, the model tends to either easily overfit to the dataset or cannot fully capture the detailed voxel information. Thus, we introduced a training strategy—random mask to address the issue. In the training stage, for each batch, we randomly masked out a cube of size m ($m \times m \times m$) from both the iCT and vCT . An illustration of the random mask strategy is shown in Figure 4.

Since the voxels in the masked cube were completely blocked out from the network in the given batch, the network would return higher losses for the masked cube, which would in turn make the model assign higher weights for the masked cube and yield images with better fine details in the next batch to reduce the loss. With the batches going on, the model would go through all the voxels and eventually result in good deformation for the entire image dataset. Moreover, the introduced random mask can also be considered as a way of data augmentation by inserting various noises (i.e., the masked cubic) for each batch, thus alleviating the risk of overfitting. Through an empirical study, we found that a random mask with a size of 5 balanced the accuracy and time cost the best, thus we used a size of 5 in the following experiments. Adam optimizer with an initial learning rate of $1e-4$ was used for training and the hyperparameters associated with the Adam optimizer were $\beta_1 = 0.9$ and $\beta_2 = 0.999$. The models were implemented with

ALGORITHM 1 The training protocol of the proposed DIR method.

INPUT Pairs of initial CTs (iCT s) and verification CTs (vCT s), number of training epochs N , learning rate η , loss balance weights α and β , mask size m , the bitmaps of the CTV, and five OAR structures s .

OUTPUT The synthetic CT (sCT), propagated contours, and the corresponding deformable vector fields (DVF).

FOR $n = 1$ to N **do**

Randomly cropping a cubic of $m \times m \times m$ from both iCT and vCT

Training the model with an objective function described in Equation (3) or (5)

Adjusting the learning rate as $\eta = \frac{\eta}{10}$ if $N \bmod 50$ equals 0.

END

RETURN sCT , propagated contours (bitmaps), and DVFs.

the PyTorch (<https://pytorch.org/>) deep learning library and the model were trained on four A100 GPUs with a batch size of 4. The proposed iterative training loop is summarized in Algorithm 1.

In the validation stage, we would not apply the random mask to the given test pairs of lung CT images. The model would produce the DVFs given the test CT images pair and then generate the sCT as mentioned before. Moreover, the bitmaps of the CTV and any given OARs contours in sCT images were generated by warping the corresponding bitmaps of structure contours from vCT images based on the generated DVFs. The details of the validation steps are shown in Algorithm 2.

2.4 | Data analysis

Both the trained wMAE and M+S models were validated in the testing dataset, which comprised of 10 independent patients and were applied with the same data pre-processing and data post-processing mentioned before. For image quality evaluation in the testing

ALGORITHM 2 The validation protocol of the proposed DIR method.

INPUT	Pairs of initial CTs (<i>iCTs</i>) and verification CTs (<i>vCTs</i>) in the testing dataset, bitmaps of ITV/PTV, five OARs structures <i>s</i> and the trained model <i>M</i> .
OUTPUT	The synthetic CT (<i>sCT</i>), propagated contours (bitmaps), and the corresponding deformable vector fields (DVF).
FOR	ANY <i>iCT</i> and <i>vCTs</i> in the testing dataset do Inputting them to the trained model <i>M</i>
END	
RETURN	<i>sCT</i> , propagated contours (bitmaps), and DVFs.

dataset, we directly measured the similarity using Per-voxel absolute CT-number-difference volume histogram (CDVH) and MAE as the evaluation metrics between the ground-truth CTs, that is, *iCT*, and the synthetic CTs (*sCT*), which were derived by deforming the *vCT* with the derived DVFs.

Four conventional DIR approaches (fast symmetric force, diffeomorphic, log domain diffeomorphic, symmetric log domain diffeomorphic) were also compared. For the conventional DIR methods, we used the DIR algorithms included in the open-source image registration library, Plastimatch,⁹² to register *vCT* to *iCT*. We used the same pre-processing procedure for the conventional DIR approaches for fair comparison.

For the dosimetric evaluation, we postprocessed both the *sCT* and *iCT* by inverting all the steps in the pre-processing stage, so that all *sCTs* and *iCTs* had the same configurations as their corresponding *iCTs*. Then, forward dose calculations of the original plan were done based on *sCT* and *iCT*. The resulting dose distributions were compared using the 3D Gamma analysis. Dose volume histograms (DVHs) were generated as well for these two dose distributions. We also compared the clinically relevant DVH indices for the selected structures. We considered D_{95} and D_5 (the minimum dose covering the highest irradiated 95% and 5% of the structure's volume, respectively) for CTV, V_5 (the minimum volume percentage receiving at least 5 Gy [RBE]) for total lung, D_{mean} (mean dose) for heart, D_{max} (max dose) for cord, and D_{mean} for esophagus. The clinically relevant DVH indices were also statistically analyzed using the paired Student's *t*-test. A *p*-value ≤ 0.05 was considered to be statistically significant.

3 | RESULTS

Sections 3.1 and 3.2 report the evaluation regarding the image quality of the *sCT*. Sections 3.3 and 3.4 show the dosimetric evaluation on the *sCT*.

TABLE 1 Comparison of the MAE and time cost of the proposed approaches and four conventional DIR approaches.

	MAE (HU)	Time cost (Seconds)
wMAE	13.15 \pm 3.8	(263.7 \pm 163) $\times 10^{-3}$
M+S	17.52 \pm 5.8	(265.8 \pm 190) $\times 10^{-3}$
FSF	56.4 \pm 18.1	280.3 \pm 129.8
DM	100.4 \pm 25.5	283.5 \pm 125.2
LD	249.0 \pm 59.4	290.9 \pm 101.5
SLD	400.96 \pm 69.4	304.3 \pm 97.2

In each cell, we reported the mean and standard deviation value of 10 test patients. Abbreviations: DM, diffeomorphic; FSF, fast symmetric force; LD, log domain diffeomorphic; SLD, symmetric log domain diffeomorphic.

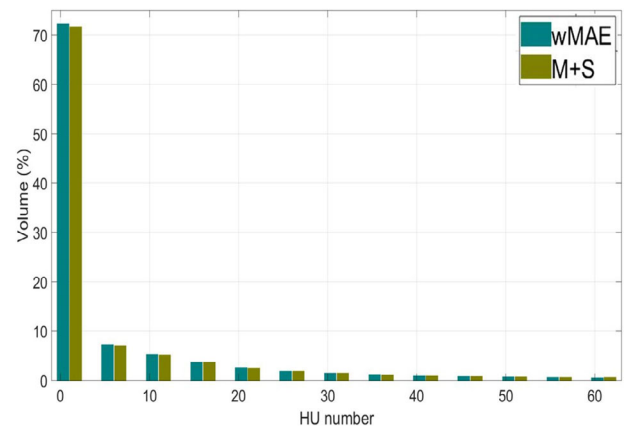


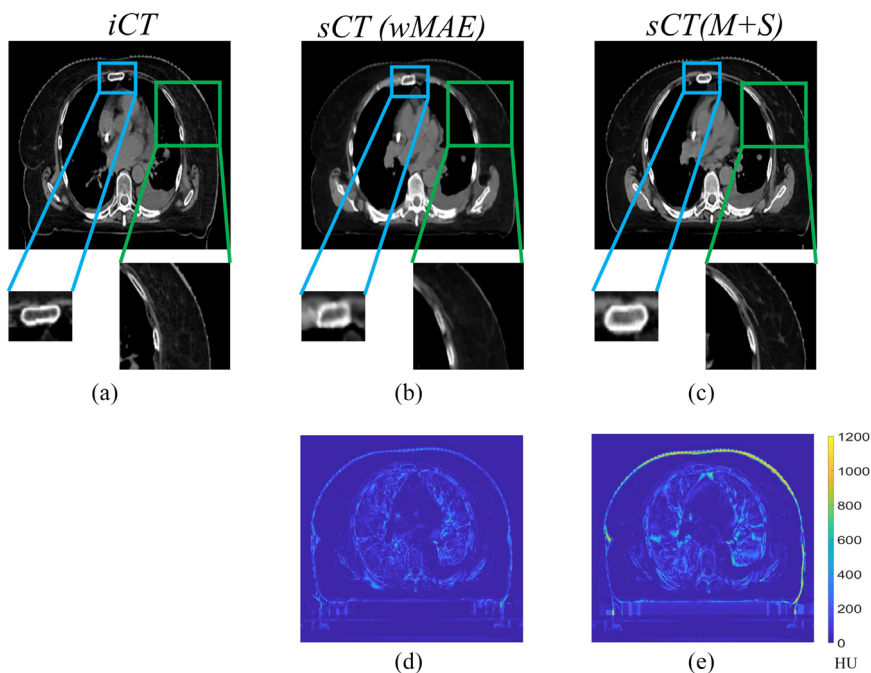
FIGURE 5 The absolute per-voxel CT number difference volume histogram for a typical patient. The x-axis represents the HU number absolute difference between the *sCT* and *iCT* and the y-axis represents the percentage of the volume. For both models, more than 70% of the volume are exactly reconstructed.

3.1 | Evaluation of the *sCT* quality

Table 1 displays the comparison of the HU number MAE and time cost of the proposed approach and four conventional DIR approaches (fast symmetric force, diffeomorphic, log domain diffeomorphic, symmetric log domain diffeomorphic). From Table 1, we observed that the proposed methods achieved better quality *sCT* as indicated by much smaller MAE compared to the conventional methods with a time cost of fewer than 300 ms, whereas all conventional DIR approaches suffered from worse quality *sCT* (as indicated by larger MAE) and all with a much longer (at least 1000 times larger than our methods) computation time.

Furthermore, we derived the per-voxel CT-number absolute difference volume histogram (CDVH) with the absolute CT number differences (in HU) as the horizontal axis and the normalized volume (in %) as the vertical axis (Figure 5) to show the per-voxel absolute CT-number difference between *sCTs* and *iCT* statistically. As shown in the figure, a majority of the voxels

FIGURE 6 Comparison of a typical slice between the *iCT* slice (a) and its corresponding *sCT* generated by the model trained with wMAE only (wMAE) (b) and the model trained with both wMAE and SSIM (M+S) (c), respectively. The CT HU number display window level and width were -125 and 1300 HU respectively. The edge of the rectum was more blurred in *sCT* generated by the wMAE model. The high Z material in *sCT* generated by the M+S model were partially recovered, whereas the wMAE model did not. Figure 5(d) and (e) showed the absolute difference of the slice between the *iCT* and *sCT* generated by the wMAE and (M+S) models, respectively.



had a small per-voxel absolute CT number difference (close to 0) between the *iCT* and *sCT*. Statistically, only 5% of the voxels have a per-voxel absolute CT number difference larger than 46.7538 HU for the model trained with the weighted MAE only and 54.6117 HU for the model trained with both the weighted MAE and SSIM, respectively.

Figure 6 compares a typical CT slice between the *iCT* (a) and the corresponding *sCT* generated by the wMAE model (b) and the M+S model (c), respectively. The differences between the *iCT* slice and *sCT* slices are shown in Figure 5d and e, where the brighter the color is, the greater the difference is. Overall, both *sCT* matched the *iCT* well and no obvious artifacts were identified. There were, however, some discrepancies between the *sCT* and the *iCT* in some soft tissue regions. When comparing the *sCT* (wMAE) and the *sCT* (M+S), the *sCT* (wMAE) achieved a lower average MAE and had much smoother edges (skin), although its image quality appeared slightly worse than the *sCT* (M+S) since the sternum was blurred (blue rectangle in Figure 6) and the high Z material in muscle was barely recovered (green rectangle in Figure 6).

3.2 | Evaluation of the propagated contours

Similar to the procedure in Section 3.1, we generated a new set of contours by propagating the contours from the *vCT* to *iCT* based on the derived DVFs from both trained models. Then we measured the similarity between the propagated contours and the initial con-

TABLE 2 Comparison of SSIMs of the selected structures generated by the models trained with wMAE and M+S, respectively.

	wMAE	M+S
CTV	0.989 ± 0.013	0.993 ± 0.009
Right lung	0.973 ± 0.011	0.975 ± 0.008
Left lung	0.976 ± 0.008	0.977 ± 0.007
Esophagus	0.997 ± 0.001	0.997 ± 0.001
Heart	0.987 ± 0.005	0.989 ± 0.005
Cord	0.998 ± 0.001	0.998 ± 0.001
average	0.987 ± 0.006	0.988 ± 0.004

The higher the SSIM value is, the higher the agreement of the selected structures between the *sCT* and *iCT* is.

tours using SSIM. The detailed results are shown in Table 2.

From Table 2, it is obvious that the proposed methods can successfully generate contours with excellent agreement for the selected structures (CTV, right lung, left lung, esophagus, heart, and cord) with the ground-truth contours after propagation (the average SSIM scores of 0.987 ± 0.006 and 0.988 ± 0.004 for the two proposed models, respectively). The model trained with M+S achieved higher SSIM scores for all structures than those of the model trained with wMAE. We also calculated the DSCs, which only consider the overlap between the ground-truth and propagated structures, for the selected structures. We found that both left and right lung suffered from low DSC scores (approximately 0.75) due to inter-fractional anatomy changes and irregular respiratory patterns, whereas the SSIM score was not greatly affected by the anatomy changes.

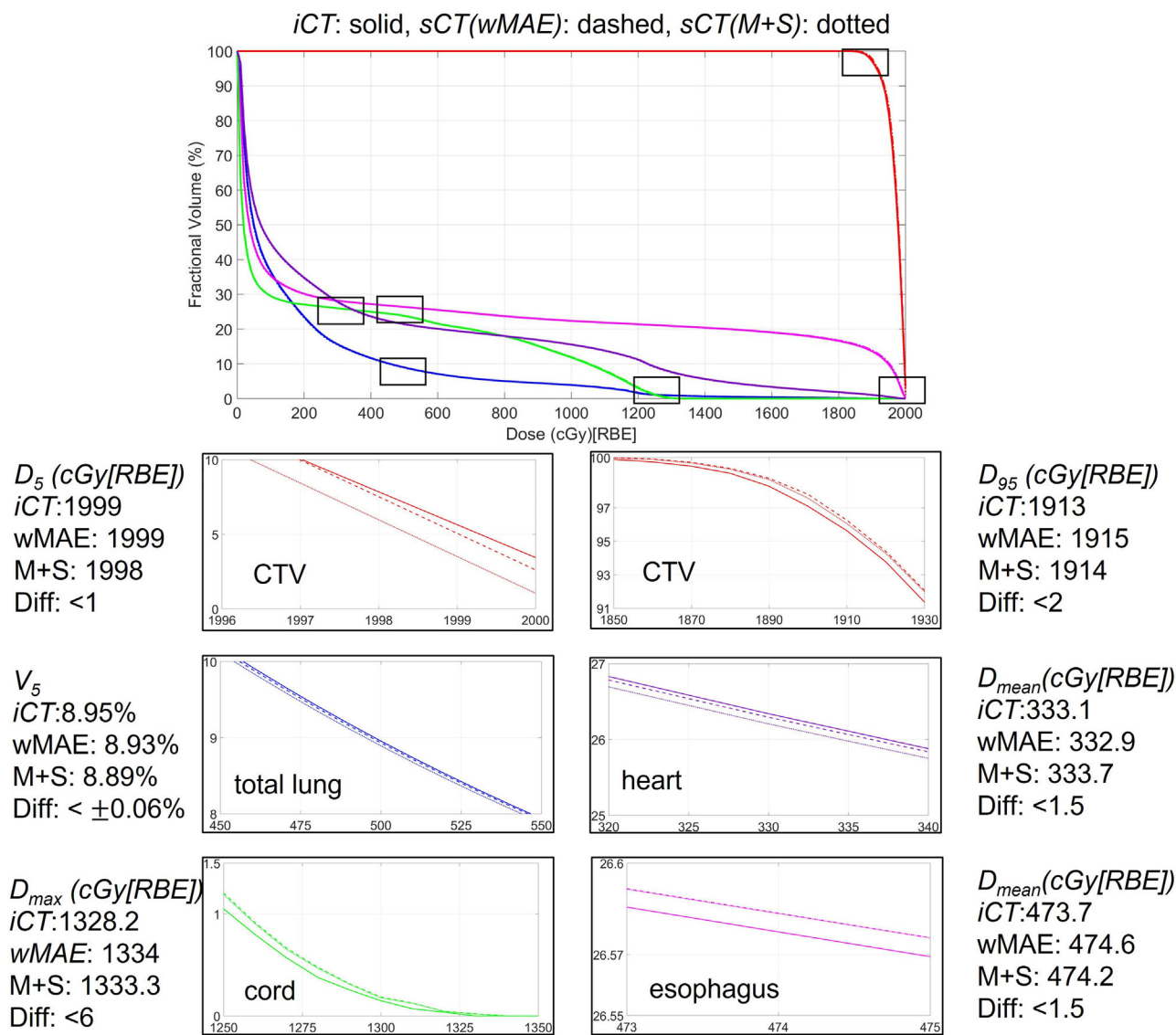


FIGURE 7 Comparison of dose volume histograms (DVHs) of one typical patient derived from the dose distributions calculated on *iCT* and the corresponding *sCT*s derived from the two models proposed in this study. In each figure, the solid, dashed, and dotted lines represent the DVHs generated from the dose distributions calculated based on *iCT*, *sCT*(wMAE) and *sCT*(M+S), respectively.

3.3 | Comparison of the dose volume histograms (DVHs) and the clinically relevant DVH indices

We performed the forward dose calculation of the original plan on the *iCT* and *sCT*. We then generated dose volume histograms (DVHs) based on the two dose distributions for every testing patient. Figure 7 shows the comparison of the DVHs generated from the dose distributions calculated on the *iCT* and its corresponding *sCT* for a typical photon patient, where the red curve represents the CTV, the blue curve represents the total lung, the purple curve represents the heart, the green curve represents the cord and the magenta curve represents the esophagus. The solid, dashed, and

dotted represent the DVHs generated from the dose distributions calculated based on *iCT*, *sCT*(wMAE) and *sCT*(M+S), respectively. For better visualization, zoom-in detailed subfigures and DVH indices differences were also provided. Visually, the DVH curves on *iCT* and *sCT*s completely overlapped with each other with negligible differences only visible in the zoomed-in regions.

Figure 8 shows the comparison of the boxplots of the clinically relevant DVH indices of the 10 testing patients from the dose distributions calculated on *iCT* and the corresponding *sCT*s derived from the two models proposed in this study. *P*-values are shown on the top of the boxplots. From Figure 8, it is clear that the clinically relevant DVH indices derived from the dose distributions calculated on the *sCT*s were very similar to the

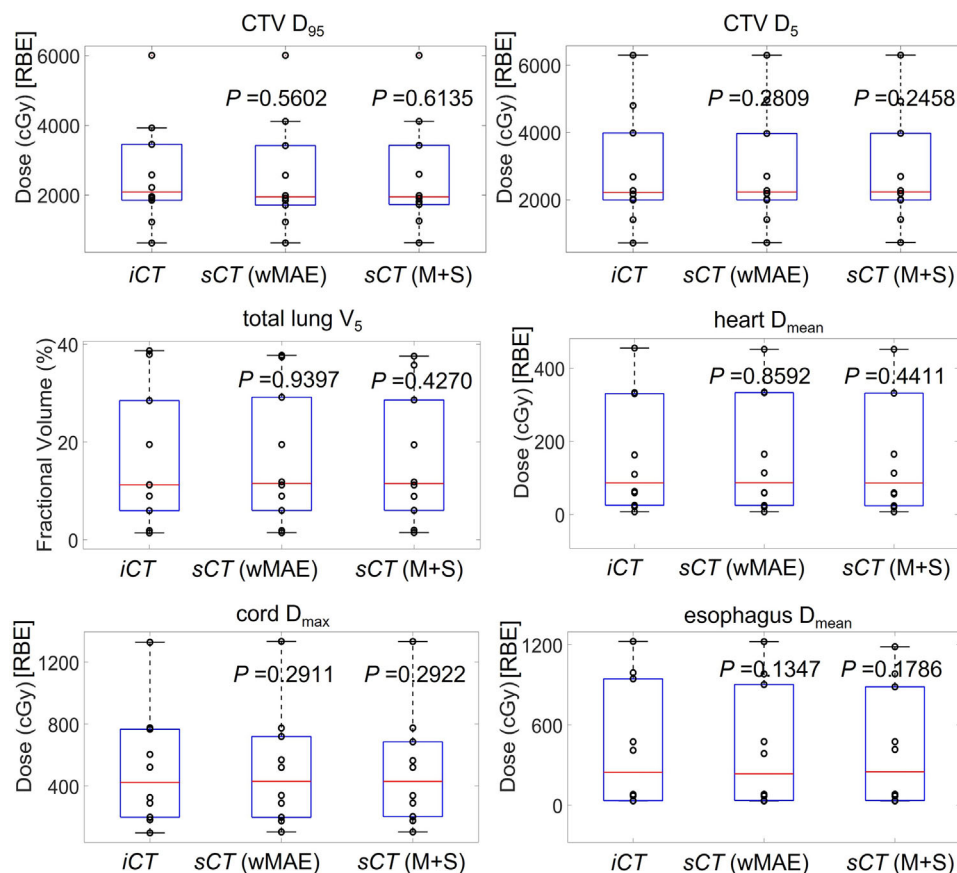


FIGURE 8 Comparison of the boxplots of the clinically relevant DVH indices of the ten testing patients from the dose distributions calculated on *iCT* and the corresponding *sCT*s derived from the two models (wMAE and M+S) proposed in this study. *P*-values derived from the statistic tests between the DVH indices calculated based on the dose distributions on *iCT* and *sCT* (wMAE), and between the DVH indices calculated based on the dose distributions on *iCT* and *sCT* (M+S) are shown on the top of the corresponding boxplots.

ones from the dose distribution calculated on *iCT* for all selected structures.

3.4 | Comparison of the dose distributions using 3D gamma analysis

We also compared the dose distributions calculated on *iCT* and the corresponding *sCT*s derived from the two models proposed in this study using 3D Gamma analysis with a threshold of 3%/3 mm/10% and 2%/2 mm/10%, respectively (Table 3). The average 3D gamma passing rate for a threshold of 3%/3 mm/10% was above 98% and above 96% for the wMAE model and the M+S model, respectively. For a threshold of 2%/2 mm/10%, the average 3D Gamma passing rate was above 97% and above 94% for the wMAE model and the M+S model, respectively. If the 10th testing patient was excluded, the average 3D Gamma passing rate for the remaining nine testing cases was above 99% for a threshold of 3%/3 mm/10% and above 98% for a threshold of 2%/2 mm/10% for the wMAE model. Moreover, the average 3D Gamma passing rate for patients treated with photon therapy was approximately

3% higher than that of patients treated with proton therapy. We also noticed that the wMAE model obtained a higher 3D Gamma passing rate than that of the M+S model.

3.5 | Ablation study

Figure 9 showed the *iCT* and corresponding *sCT*s generated by different models trained with different loss terms. Figure 9a and b show the *sCT*s generated by the model trained with the SSIM or the MSE loss term only, respectively. Figure 9b was the model setting adopted by the original Voxelmorph model. Figure 9c presented the result generated by the model trained with both the MSE and SSIM loss terms, Figure 9d was for the model trained with the wMAE loss term without applying random mask strategy and Figure 9e was the *iCT*. Figure 9f-i were the absolute difference between the *sCT*s and *iCT*. Compared with the results showed in Figure 6, it was clearly seen that no models shown in Figure 9 generated the *sCT*s with better details, for example, neither of the models can generate *sCT* with corrected rectum nor cord. We further quantitatively

TABLE 3 The 3D Gamma passing rates between the dose distributions calculated on *iCT* and the corresponding *sCT*s of the 10 testing patients with a threshold of 3%/3 mm/10% and 2%/2 mm/10% for both the wMAE model and the M+S model, respectively.

PATIENT	3%/3 mm/10% (wMAE) / (M+S)	2%/2 mm/10%(wMAE) / (M+S)
#1	0.999 / 0.998	0.997 / 0.996
#2 ^a	0.996 / 0.965	0.98 / 0.926
#3	1.0 / 0.991	0.999 / 0.977
#4	1.0 / 0.999	0.999 / 0.999
#5	0.995 / 0.943	0.978 / 0.913
#6 ^a	0.966 / 0.976	0.921 / 0.934
#7	1.0 / 0.979	1.0 / 0.970
#8	1.0 / 0.979	1.0 / 0.970
#9	0.999 / 0.924	0.991 / 0.811
#10 ^a	0.921 / 0.924	0.899 / 0.902
Average	0.986±0.026 / 0.963±0.029	0.977±0.036 / 0.945±0.062
Average(photon)	0.995±0.012 / 0.974±0.029	0.987±0.029 / 0.955±0.067
Average(proton)	0.971±0.043 / 0.944±0.020	0.953±0.047 / 0.907±0.007

^aThe patient was treated with proton therapy.

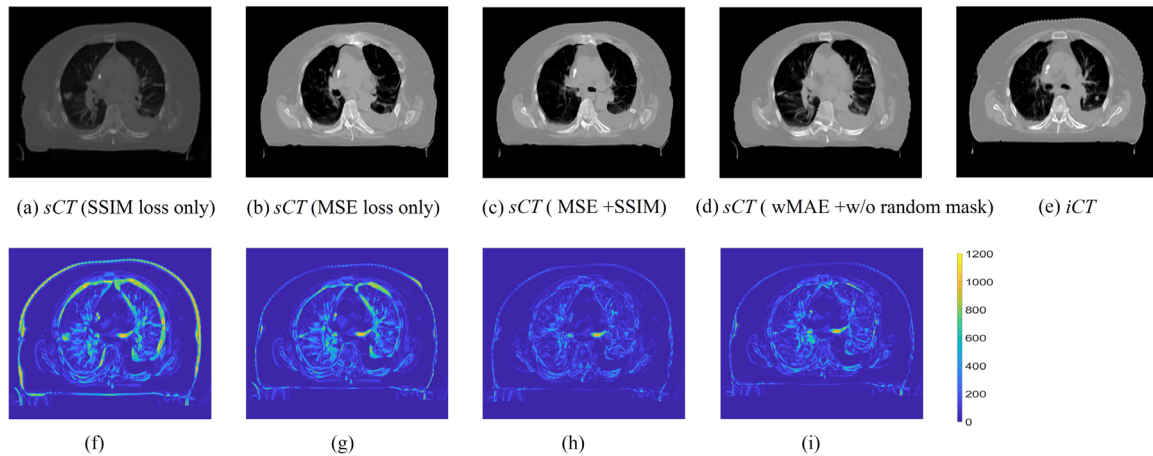


FIGURE 9 Sample slices of the *sCT*s generated by the models trained with different loss terms and the corresponding *iCT* (e). Figure (a) and (b) are the sample slices of the *sCT*s generated by the models trained with the SSIM or MSE loss term only, respectively. Figure (c) shows the sample slice of the *sCT* generated by the model trained with both the MSE and SSIM loss terms. Figure (d) shows the sample slice of the *sCT* generated by the model trained with the wMAE loss term without random mask strategy. The CT HU number display window position and width were -120 and 1300 HU, respectively. Figure (f)-(i) show the corresponding absolute differences between the *sCT*s and *iCT*.

TABLE 4 The MAE of the *sCT* generated by models trained with different loss terms.

	MAE(HU)
SSIM loss only	52.73
MSE loss only	26.12
MSE + SSIM	19.42
wMAE w/o random mask	22.64

evaluated the quality of the *sCT*s by different models. The *sCT*s shown in Figure 9a,b,c and d achieved a MAE of 52.73, 26.12, 19.42, and 22.64 HU, respectively (Table 4), while the wMAE and M+S models proposed

in this work achieved a MAE of 13.15 and 17.52HU, respectively (Table 1).

4 | DISCUSSION

In this study, we have developed a VoxelMorph-based deep neural network for fast and accurate DIR in the radiotherapy of lung cancer. We tried two configurations (thus two models) in the proposed methods and performed a comprehensive validation of the proposed models on the CT images of lung cancer patients. Although the methods based on CT images were focused on lung cancer, the methods could be generalized to all disease sites.

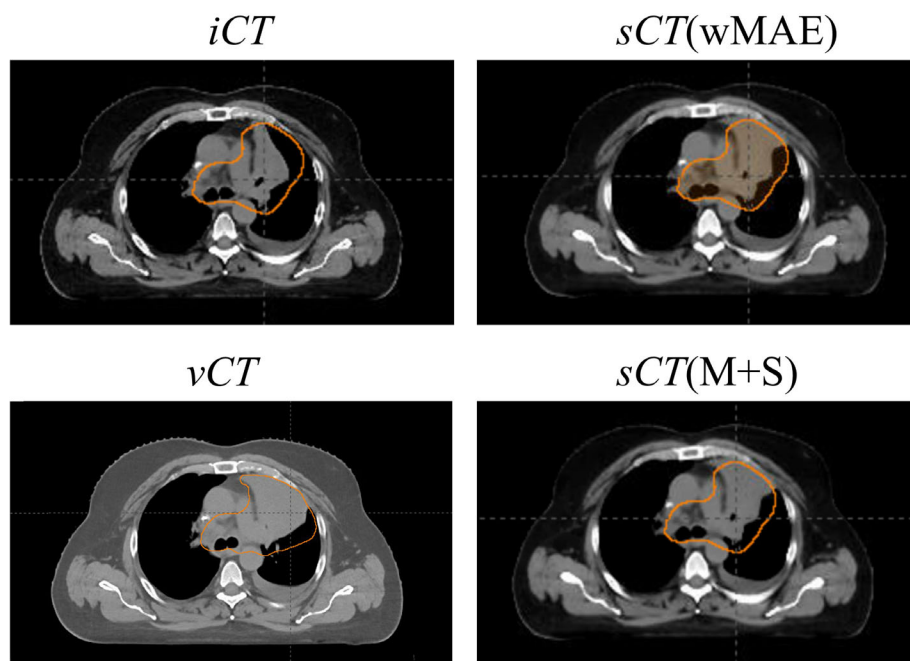


FIGURE 10 Comparison of one CT slice in the middle of CTV among the *iCT*, *vCT*, and *sCT*s of a case with a relatively poor performance from our proposed methods (the 10th testing patient in Table 3), where the circle highlights the CTV in each subfigure.

To alleviate the potential overfitting caused by limited data and low resolution, we introduced a random mask training strategy, and included additional loss terms in the objective function (i.e., the weighted MAE and SSIM terms), to improve the quality of the *sCT*. Through an empirical study, we found that a random mask with a size of $5 \times 5 \times 5$ yielded the optimal performance in our study. A random mask with a very small size cannot mitigate the blur in the *sCT* images well enough, whereas a random mask with a very large size, though it may help to mitigate the blur, introduces uncertainty to the model training and eventually slows down or even collapses the training of the neural network. As for the loss terms, we used the weighted MAE to guarantee the voxel-to-voxel similarity, in which the per-voxel loss weight is proportional to the HU numbers, thus helping to reduce the high-frequency artifacts (e.g., the artifacts in bone structure). Comparing with the MSE loss term, which is used by the original Voxelmorph model, the use of the wMAE loss term greatly improves the quality of the *sCT*. Another additional loss term used in this study is the SSIM, which is a loss that enforces the similarity among structures. However, unlike dice similarity coefficients (DSCs),^{22,90,91} which has been extensively used as the structure similarity evaluation metric, SSIM considers not only the similarity among structures, but also the illumination and contrast among the images. Thus, it is a better choice for the lung CT images since the lung CT images often have multiple structures with large variation of the HU numbers. This potentially leads to diverse illuminations and contrasts.

We trained the proposed neural network with two configurations (thus two models), one with the weighted

MAE loss term only (wMAE model) and the other with both the weighted MAE loss term and SSIM loss term (M+S model). The results related to *sCT* quality during the validation showed that the M+S model yielded slightly better *sCT* image quality than that of the wMAE model. However, the dosimetric evaluation by comparing the clinically relevant DVH indices and performing 3D Gamma analysis between the dose distributions calculated on *iCT* and the corresponding *sCT*s derived from the two models proposed in this study presented the opposite results—the wMAE model had a slightly better agreement of the DVH indices with ground-truth and a higher 3D Gamma passing rate. This indicates that the evaluation of the *sCT* quality needs to be conducted thoroughly and cannot rely on one criterion alone. Additionally, it further suggests that high *sCT* image quality, although clinically relevant in radiation therapy, does not necessarily lead to favorable results in dose calculation. How to further improve the performance of the proposed deep neural network in dose calculation will be an interesting and challenging research direction.

The 3D Gamma passing rates reported in Table 3 are very promising and exceeds the clinical requirements suggested by American Association of Physicists in Medicine (AAPM) task group (TG) 218,⁹³ suggesting that the generated *sCT*s can be reliably used in clinical applications such as ART. However, it requires further improvements. There are multiple factors for these non-ideal results. One factor is inherited in the different physics characteristics of proton and photon therapy, where proton therapy could be more sensitive to the same variations in the HU numbers compared to

photon therapy as shown in Table 3. Another possible contributing factor can be the challenging CT imaging dataset used in this study: all the CT images pairs (*iCT* and *vCT*) used in this study are the CT images taken at least several weeks apart, during which the inter-fractional anatomy changes can be large, irregular, and unpredictable. In addition, the tumor may grow or shrink and patients' weight may change during the time window between the *iCT* and *vCT*, thus introducing additional unpredictable ambiguities (new information or loss of the old information compared to the information contained in *iCT*) for DIR. Figure 10 shows the comparison of one CT slice in the middle of the tumor among the *iCT*, *vCT* and *sCTs* of a case with a relatively poor performance from our proposed methods (the 10th testing patient in Table 3), where the circle highlights the CTV in each subfigure. It is obvious that the generated *sCTs* have a CTV with larger high-density regions compared to the ground-truth CT (i.e., *iCT*), possibly due to the fact that this patient has a very aggressive tumor phenotype (the tumor grows a lot from *iCT* to *vCT*). This might lead to worse agreements of DVH indices and a lower 3D Gamma passing rate (92.11% of 3 mm/3%/10% and 89.92% of 2 mm/2%/10% for the wMAE model and 92.30% of 3 mm/3%/10% and 90.21% of 2 mm/2%/10% for the M+S model, respectively). Moreover, for more challenging disease sites which have complexity shapes, such as ovarian cancer⁹⁴ and for registration between different modalities,⁹⁵ the proposed DIR approach may see its limitation. Further investigation is needed to address these issues.

5 | CONCLUSION

A deep neural network-based DIR approach was proposed and shown to be accurate and efficient to register the initial CTs and verification CTs for lung cancer.

ACKNOWLEDGMENTS

This research was supported by the National Cancer Institute (NCI) Career Developmental Award K25CA168984, Arizona Biomedical Research Commission Investigator Award, the President's Discovery and Translational Program of Mayo Clinic, the Fred C. and Katherine B. Andersen Foundation Translational Cancer Research Award, the Lawrence W. and Marilyn W. Matteson Fund for Cancer Research, and the Kemper Marley Foundation.

CONFLICT OF INTEREST STATEMENT

Terence T. Sio provides strategic and scientific recommendations as a member of the Advisory Board and speaker for Novocure, Inc., Catalyst Pharmaceuticals, Inc. and Galera Pharmaceuticals, which are not in

any way associated with the content presented in this manuscript.

REFERENCES

- Oh S, Kim S. Deformable image registration in radiation therapy. *Radiat Oncol J*. 2017;35(2):101-111. [published online ahead of print 2017/07/18].
- Thor M, Petersen JB, Bentzen L, Hoyer M, Muren LP. Deformable image registration for contour propagation from CT to cone-beam CT scans in radiotherapy of prostate cancer. *Acta Oncol (Stockholm, Sweden)*. 2011;50(6):918-925. [published online ahead of print 2011/07/20].
- Hautvast G, Lobregt S, Breeuwer M, Gerritsen F. Automatic contour propagation in cine cardiac magnetic resonance images. *IEEE Trans Med Imaging*. 2006;25(11):1472-1482. [published online ahead of print 2006/11/23].
- Söhn M, Birkner M, Yan D, Alber M. Modelling individual geometric variation based on dominant eigenmodes of organ deformation: implementation and evaluation. *Phys Med Biol*. 2005;50(24):5893-5908. [published online ahead of print 2005/12/08].
- Nguyen TN, Moseley JL, Dawson LA, Jaffray DA, Brock KK. Adapting liver motion models using a navigator channel technique. *Med Phys*. 2009;36(4):1061-1073. [published online ahead of print 2009/05/29].
- Budiarto E, Keijzer M, Storchi PR, et al. A population-based model to describe geometrical uncertainties in radiotherapy: applied to prostate cases. *Phys Med Biol*. 2011;56(4):1045-1061. [published online ahead of print 2011/01/25].
- Oh S, Jaffray D, Cho YB. A novel method to quantify and compare anatomical shape: application in cervix cancer radiotherapy. *Phys Med Biol*. 2014;59(11):2687-2704. [published online ahead of print 2014/05/03].
- Guerrero T, Sanders K, Noyola-Martinez J, et al. Quantification of regional ventilation from treatment planning CT. *Int J Radiat Oncol Biol Phys*. 2005;62(3):630-634. [published online ahead of print 2005/06/07].
- Yaremko BP, Guerrero TM, Noyola-Martinez J, et al. Reduction of normal lung irradiation in locally advanced non-small-cell lung cancer patients, using ventilation images for functional avoidance. *Int J Radiat Oncol Biol Phys*. 2007;68(2):562-571. [published online ahead of print 2007/04/03].
- Yamamoto T, Kabus S, von Berg J, Lorenz C, Keall PJ. Impact of four-dimensional computed tomography pulmonary ventilation imaging-based functional avoidance for lung cancer radiotherapy. *Int J Radiat Oncol Biol Phys*. 2011;79(1):279-288. [published online ahead of print 2010/07/22].
- Qi XS, Santhanam A, Neylon J, et al. Near real-time assessment of anatomic and dosimetric variations for head and neck radiation therapy via graphics processing unit-based dose deformation framework. *Int J Radiat Oncol Biol Phys*. 2015;92(2):415-422. [published online ahead of print 2015/04/08].
- Sharma M, Weiss E, Siebers JV. Dose deformation-invariance in adaptive prostate radiation therapy: implication for treatment simulations. *Radiation Oncol*. 2012;105(2):207-213. [published online ahead of print 2012/12/04].
- Velec M, Moseley JL, Eccles CL, et al. Effect of breathing motion on radiotherapy dose accumulation in the abdomen using deformable registration. *Int J Radiat Oncol Biol Phys*. 2011;80(1):265-272. [published online ahead of print 2010/08/25].
- Yan D, Vicini F, Wong J, Martinez A. Adaptive radiation therapy. *Phys Med Biol*. 1997;42(1):123-132. [published online ahead of print 1997/01/01].
- Schalj B, Kempe JA, Bauman GS, Battista JJ, Van Dyk J. Tracking the dose distribution in radiation therapy by accounting for variable anatomy. *Phys Med Biol*. 2004;49(5):791-805. [published online ahead of print 2004/04/09].

16. Christensen GE, Carlson B, Chao KS, et al. Image-based dose planning of intracavitary brachytherapy: registration of serial-imaging studies using deformable anatomic templates. *Int J Radiat Oncol Biol Phys*. 2001;51(1):227-243. [published online ahead of print 2001/08/23].
17. Yan D, Jaffray DA, Wong JW. A model to accumulate fractionated dose in a deforming organ. *Int J Radiat Oncol Biol Phys*. 1999;44(3):665-675. [published online ahead of print 1999/05/29].
18. Vercauteren T, Pennec X, Perchant A, Ayache N. Non-parametric diffeomorphic image registration with the demons algorithm. Paper presented at International Conference on Medical Image Computing and Computer-Assisted Intervention 2007.
19. Zhong H, Kim J, Li H, Nurushev T, Movsas B, Chetty IJ. A finite element method to correct deformable image registration errors in low-contrast regions. *Phys Med Biol*. 2012;57(11):3499-3515. [published online ahead of print 2012/05/15].
20. Gu X, Dong B, Wang J, et al. A contour-guided deformable image registration algorithm for adaptive radiotherapy. *Phys Med Biol*. 2013;58(6):1889-1901. [published online ahead of print 2013/02/28].
21. Nithiananthan S, Schafer S, Mirota DJ, et al. Extra-dimensional Demons: a method for incorporating missing tissue in deformable image registration. *Med Phys*. 2012;39(9):5718-5731. [published online ahead of print 2012/09/11].
22. Lawson JD, Schreibmann E, Jani AB, Fox T. Quantitative evaluation of a cone-beam computed tomography-planning computed tomography deformable image registration method for adaptive radiation therapy. *J Appl Clin Med Phys*. 2007;8(4):96-113. [published online ahead of print 2008/05/02].
23. Kierkels RGJ, den Otter LA, Korevaar EW, et al. An automated, quantitative, and case-specific evaluation of deformable image registration in computed tomography images. *Phys Med Biol*. 2018;63(4):045026. [published online ahead of print 2017/11/29].
24. Mencarelli A, van Kranen SR, Hamming-Vrieze O, et al. Deformable image registration for adaptive radiation therapy of head and neck cancer: accuracy and precision in the presence of tumor changes. *Int J Radiat Oncol Biol Phys*. 2014;90(3):680-687. [published online ahead of print 2014/08/26].
25. Jacobson TJ, Murphy MJ. Optimized knot placement for B-splines in deformable image registration. *Med Phys*. 2011;38(8):4579-4582. [published online ahead of print 2011/09/21].
26. Shekhar R, Lei P, Castro-Pareja CR, Plishker WL, D'Souza WD. Automatic segmentation of phase-correlated CT scans through nonrigid image registration using geometrically regularized free-form deformation. *Med Phys*. 2007;34(7):3054-3066.
27. Reed VK, Woodward WA, Zhang LF, et al. Automatic segmentation of whole breast using atlas approach and deformable image registration. *Int J Radiat Oncol*. 2009;73(5):1493-1500.
28. Wang H, Dong L, O'Daniel J, et al. Validation of an accelerated 'demons' algorithm for deformable image registration in radiation therapy. *Phys Med Biol*. 2005;50(12):2887-2905.
29. Nocedal J. Updating quasi-Newton matrices with limited storage. *Math Comput*. 1980;35(151):773-782.
30. Ong CL, Verbakel WF, Cuijpers JP, Slotman BJ, Lagerwaard FJ, Senan S. Stereotactic radiotherapy for peripheral lung tumors: a comparison of volumetric modulated arc therapy with 3 other delivery techniques. *Radiother Oncol*. 2010;97(3):437-442.
31. McGrath SD, Matuszak MM, Yan D, Kestin LL, Martinez AA, Grills IS. Volumetric modulated arc therapy for delivery of hypofractionated stereotactic lung radiotherapy: a dosimetric and treatment efficiency analysis. *Radiother Oncol*. 2010;95(2):153-157.
32. Chan OS, Lee MC, Hung AW, Chang AT, Yeung RM, Lee AW. The superiority of hybrid-volumetric arc therapy (VMAT) technique over double arcs VMAT and 3D-conformal technique in the treatment of locally advanced non-small cell lung cancer—A planning study. *Radiother Oncol*. 2011;101(2):298-302.
33. Chun SG, Hu C, Choy H, et al. Impact of intensity-modulated radiation therapy technique for locally advanced non-small-cell lung cancer: a secondary analysis of the NRG Oncology RTOG 0617 randomized clinical trial. *J Clin Oncol*. 2017;35(1):56-62.
34. Liu W, Patel SH, Shen JJ, et al. Robustness quantification methods comparison in volumetric modulated arc therapy to treat head and neck cancer. *Pract Radiat Oncol*. 2016;6(6):e269-e275.
35. van de Water TA, Bijl HP, Schilstra C, Pijs-Johannesma M, Langendijk JA. The potential benefit of radiotherapy with protons in head and neck cancer with respect to normal tissue sparing: a systematic review of literature. *Oncologist*. 2011;16(3):366-377. [published online ahead of print 2011/02/26].
36. Lin A, Swisher-McClure S, Millar LB, et al. Proton therapy for head and neck cancer: current applications and future directions. *Transl Cancer Res*. 2012;1(4):255-263.
37. Frank SJ, Cox JD, Gillin M, et al. Multifield optimization intensity modulated proton therapy for head and neck tumors: a translation to practice. *Int J Radiat Oncol Biol Phys*. 2014;89(4):846-853. [published online ahead of print 2014/05/29].
38. Schild SE, Rule WG, Ashman JB, et al. Proton beam therapy for locally advanced lung cancer: a review. *World J Clin Oncol*. 2014;5(4):568-575. [published online ahead of print 2014/10/11].
39. Bhangoo RS, DeWees TA, Yu NY, et al. Acute toxicities and short-term patient outcomes after intensity-modulated proton beam radiation therapy or intensity-modulated photon radiation therapy for esophageal carcinoma: a Mayo Clinic experience. *Adv Radiat Oncol*. 2020;5(5):871-879. [published online ahead of print 2020/10/22].
40. Bhangoo RS, Mullikin TC, Ashman JB, et al. Intensity modulated proton therapy for hepatocellular carcinoma: initial clinical experience. *Adv Radiat Oncol*. 2021;6(4):100675. [published online ahead of print 2021/08/20].
41. Yang Y, Muller OM, Shiraishi S, et al. Empirical relative biological effectiveness (RBE) for mandible osteoradionecrosis (ORN) in head and neck cancer patients treated with pencil-beam-scanning proton therapy (PBSPT): a retrospective, case-matched cohort study. *Front Oncol*. 2022;12:843175. [published online ahead of print 2022/03/22].
42. Yu NY, DeWees TA, Liu C, et al. Early outcomes of patients with locally advanced non-small cell lung cancer treated with intensity-modulated proton therapy versus intensity-modulated radiation therapy: the Mayo clinic experience. *Adv Radiat Oncol*. 2020;5(3):450-458. [published online ahead of print 2020/06/13].
43. Yu NY, DeWees TA, Voss MM, et al. Cardiopulmonary toxicity following intensity-modulated proton therapy (IMPT) vs. intensity-modulated radiation therapy (IMRT) for stage III non-small cell lung cancer. *Clin Lung Cancer*. 2022;23(8):e526-e535. doi:10.1016/j.clcc.2022.07.017
44. Lomax AJ. Intensity modulated proton therapy and its sensitivity to treatment uncertainties 1: the potential effects of calculational uncertainties. *Phys Med Biol*. 2008;53(4):1027-1042. [published online ahead of print 2008/02/12].
45. Lomax AJ. Intensity modulated proton therapy and its sensitivity to treatment uncertainties 2: the potential effects of inter-fraction and inter-field motions. *Phys Med Biol*. 2008;53(4):1043-1056. [published online ahead of print 2008/02/12].
46. Lomax AJ, Pedroni E, Rutz H, Goitein G. The clinical potential of intensity modulated proton therapy. *Z Med Phys*. 2004;14(3):147-152. [published online ahead of print 2004/10/07].
47. Pflugfelder D, Wilkens JJ, Oelfke U. Worst case optimization: a method to account for uncertainties in the optimization of intensity modulated proton therapy. *Phys Med Biol*. 2008;53(6):1689-1700. [published online ahead of print 2008/03/28].
48. Unkelbach J, Bortfeld T, Martin BC, Soukup M. Reducing the sensitivity of IMPT treatment plans to setup errors and range uncertainties via probabilistic treatment planning. *Med Phys*. 2009;36(1):149-163. [published online ahead of print 2009/02/25].
49. Fredriksson A, Forsgren A, Hardemark B. Minimax optimization for handling range and setup uncertainties in proton therapy. *Med*

- Phys.* 2011;38(3):1672-1684. [published online ahead of print 2011/04/28].
50. Liu W, Zhang X, Li Y, Mohan R. Robust optimization in intensity-modulated proton therapy. *Med Phys.* 2012;39:1079-1091.
 51. Liu W, Li Y, Li X, Cao W, Zhang X. Influence of robust optimization in intensity-modulated proton therapy with different dose delivery techniques. *Med Phys.* 2012;39(6):3089-3101. [published online ahead of print 2012/07/05].
 52. Chen W, Unkelbach J, Trofimov A, et al. Including robustness in multi-criteria optimization for intensity-modulated proton therapy. *Phys Med Biol.* 2012;57(3):591-608. [published online ahead of print 2012/01/10].
 53. Liu W, Liao Z, Schild SE, et al. Impact of respiratory motion on worst-case scenario optimized intensity modulated proton therapy for lung cancers. *Pract Radiat Oncol.* 2015;5(2):e77-86. [published online ahead of print 2014/11/22].
 54. Unkelbach J, Alber M, Bangert M, et al. Robust radiotherapy planning. *Phys Med Biol.* 2018;63(22):22TR02. [published online ahead of print 2018/11/13].
 55. An Y, Liang J, Schild SE, Bues M, Liu W. Robust treatment planning with conditional value at risk chance constraints in intensity-modulated proton therapy. *Med Phys.* 2017;44(1):28-36. [published online ahead of print 2017/01/04].
 56. An Y, Shan J, Patel SH, et al. Robust intensity-modulated proton therapy to reduce high linear energy transfer in organs at risk. *Med Phys.* 2017;44(12):6138-6147. [published online ahead of print 2017/10/05].
 57. Liu C, Patel SH, Shan J, et al. Robust optimization for intensity modulated proton therapy to redistribute high linear energy transfer from nearby critical organs to tumors in head and neck cancer. *Int J Radiat Oncol Biol Phys.* 2020;107(1):181-193. [published online ahead of print 2020/01/29].
 58. Liu C, Yu NY, Shan J, et al. Technical note: treatment planning system (TPS) approximations matter - comparing intensity-modulated proton therapy (IMPT) plan quality and robustness between a commercial and an in-house developed TPS for nonsmall cell lung cancer (NSCLC). *Med Phys.* 2019;46(11):4755-4762. [published online ahead of print 2019/09/10].
 59. Liu C, Schild SE, Chang JY, et al. Impact of spot size and spacing on the quality of robustly optimized intensity modulated proton therapy plans for lung cancer. *Int J Radiat Oncol Biol Phys.* 2018;101(2):479-489. [published online ahead of print 2018/03/20].
 60. Liu W. Inventor. System and Method for Robust Intensity-modulated Proton Therapy Planning. 2014.
 61. Liu W. Robustness quantification and robust optimization in intensity-modulated proton therapy. In: Rath A, Sahoo N, eds. *Particle Radiotherapy: Emerging Technology for Treatment of Cancer*. Springer; 2015;139-155.
 62. Liu W, Frank SJ, Li X, et al. Effectiveness of robust optimization in intensity-modulated proton therapy planning for head and neck cancers. *Med Phys.* 2013;40(5):051711. [published online ahead of print 2013/05/03].
 63. Liu W, Frank SJ, Li X, Li Y, Zhu RX, Mohan R. PTV-based IMPT optimization incorporating planning risk volumes vs robust optimization. *Med Phys.* 2013;40(2):021709. [published online ahead of print 2013/02/08].
 64. Liu W, Schild SE, Chang JY, et al. Exploratory study of 4D versus 3D robust optimization in intensity modulated proton therapy for lung cancer. *Int J Radiat Oncol Biol Phys.* 2016;95(1):523-533. [published online ahead of print 2016/01/05].
 65. Shan J, Sio TT, Liu C, Schild SE, Bues M, Liu W. A novel and individualized robust optimization method using normalized dose interval volume constraints (NDIVC) for intensity-modulated proton radiotherapy. *Med Phys.* 2019;46:382-393. doi:10.1002/mp.13276. [published online ahead of print 2018/11/06].
 66. Shan J, An Y, Bues M, Schild SE, Liu W. Robust optimization in IMPT using quadratic objective functions to account for the minimum MU constraint. *Med Phys.* 2018;45(1):460-469. [published online ahead of print 2017/11/18].
 67. Feng H, Shan J, Ashman JB, et al. 4D robust optimization in small spot intensity-modulated proton therapy (IMPT) for distal esophageal carcinoma. *Med Phys.* 2021;48(8):4636-4647.
 68. Feng H, Sio TT, Rule WG, et al. Beam angle comparison for distal esophageal carcinoma patients treated with intensity-modulated proton therapy. *J Appl Clin Med Phys.* 2020;21(11):141-152. [published online ahead of print 2020/10/16].
 69. Li Y, Liu W, Li X, Quan E, Zhang X. Toward a thorough evaluation of IMPT plan sensitivity to uncertainties: revisit the worst-case analysis with an exhaustively sampling approach. *Med Phys.* 2011;38(6):3853.
 70. Liu C, Bhangoo RS, Sio TT, et al. Dosimetric comparison of distal esophageal carcinoma plans for patients treated with small-spot intensity-modulated proton versus volumetric-modulated arc therapies. *J Appl Clin Med Phys.* 2019;20(7):15-27. [published online ahead of print 2019/05/22].
 71. Liu C, Sio TT, Deng W, et al. Small-spot intensity-modulated proton therapy and volumetric-modulated arc therapies for patients with locally advanced non-small-cell lung cancer: a dosimetric comparative study. *J Appl Clin Med Phys.* 2018;19(6):140-148. [published online ahead of print 2018/10/18].
 72. Tryggstad EJ, Liu W, Pepin MD, Hallemeier CL, Sio TT. Managing treatment-related uncertainties in proton beam radiotherapy for gastrointestinal cancers. *J Gastrointest Oncol.* 2020;11(1):212-224. [published online ahead of print 2020/03/17].
 73. Zaghian M, Cao W, Liu W, et al. Comparison of linear and nonlinear programming approaches for "worst case dose" and "minmax" robust optimization of intensity-modulated proton therapy dose distributions. *J Appl Clin Med Phys.* 2017;18(2):15-25. [published online ahead of print 2017/03/17].
 74. Zaghian M, Lim G, Liu W, Mohan R. An automatic approach for satisfying dose-volume constraints in linear fluence map optimization for IMPT. *J Cancer Ther.* 2014;5(2):198-207. [published online ahead of print 2014/12/17].
 75. Matney J, Park PC, Bluett J, et al. Effects of respiratory motion on passively scattered proton therapy versus intensity modulated photon therapy for stage III lung cancer: are proton plans more sensitive to breathing motion? *Int J Radiat Oncol Biol Phys.* 2013;87(3):576-582. [published online ahead of print 2013/10/01].
 76. Matney JE, Park PC, Li H, et al. Perturbation of water-equivalent thickness as a surrogate for respiratory motion in proton therapy. *J Appl Clin Med Phys.* 2016;17(2):368-378. [published online ahead of print 2016/04/14].
 77. Liu W, Patel SH, Harrington DP, et al. Exploratory study of the association of volumetric modulated arc therapy (VMAT) plan robustness with local failure in head and neck cancer. *J Appl Clin Med Phys.* 2017;18(4):76-83. [published online ahead of print 2017/05/16].
 78. Li H, Zhang X, Park P, et al. Robust optimization in intensity-modulated proton therapy to account for anatomy changes in lung cancer patients. *Radiother Oncol.* 2015;114(3):367-372. [published online ahead of print 2015/02/25].
 79. Bibault JE, Arsene-Henry A, Durdur C, et al. Adaptive radiation therapy for non-small cell lung cancer. *Cancer Radiother.* 2015;19(6-7):458-462. [published online ahead of print 2015/09/05].
 80. Yan D, Wong J, Vicini F, et al. Adaptive modification of treatment planning to minimize the deleterious effects of treatment setup errors. *Int J Radiat Oncol Biol Phys.* 1997;38(1):197-206. [published online ahead of print 1997/04/01].
 81. Yang X, Kwitt R, Styner M, Quicksilver Niethammer M. Fast predictive image registration - a deep learning approach. *Neuroimage.* 2017;158:378-396. [published online ahead of print 2017/07/15].
 82. Balakrishnan G, Zhao A, Sabuncu MR, Guttag J, Dalca AV. VoxelMorph: a learning framework for deformable medical image registration. *IEEE Trans Med Imaging.* 2019;38(8):1788-1800.

- doi:10.1109/TMI.2019.2897538. [published online ahead of print 2019/02/05].
83. de Vos BD, Berendsen FF, Viergever MA, Sokooti H, Staring M, Išgum I. A deep learning framework for unsupervised affine and deformable image registration. *Med Image Anal.* 2019;52:128-143. [published online ahead of print 2018/12/24].
 84. Zhao S, Dong Y, Chang EI, Xu Y. Recursive cascaded networks for unsupervised medical image registration. Paper presented at Proceedings of the IEEE/CVF international conference on computer vision 2019.
 85. He K, Chen X, Xie S, Li Y, Dollár P, Girshick R. Masked Autoencoders are Scalable Vision Learners. Paper presented at 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR); 18–24 June 2022, 2022.
 86. Subhas N, Primak AN, Obuchowski NA, et al. Iterative metal artifact reduction: evaluation and optimization of technique. *Skeletal Radiol.* 2014;43(12):1729-1735. [published online ahead of print 2014/08/31].
 87. Avants BB, Tustison NJ, Stauffer M, Song G, Wu BH, Gee JC, The Insight ToolKit image registration framework. *Front Neuroinform.* 2014;8.
 88. Maas AL, Hannun AY, Ng AY. Rectifier Nonlinearities Improve Neural Network Acoustic Models. Paper presented at: Proc. icml2013.
 89. Wang Z, Bovik AC, Sheikh HR, Simoncelli EP. Image quality assessment: from error visibility to structural similarity. *IEEE Trans Image Process.* 2004;13(4):600-612. [published online ahead of print 2004/09/21].
 90. Sorensen TA. A method of establishing groups of equal amplitude in plant sociology based on similarity of species content and its application to analyses of the vegetation on Danish commons. *Biol Skar.* 1948;5:1-34.
 91. Dice LR. Measures of the amount of ecologic association between species. *Ecology.* 1945;26(3):297-302.
 92. Sharp GC, Li R, Wolfgang J, et al. Plastimatch: an open source software suite for radiotherapy image processing. Paper presented at: Proceedings of the XVI'th International Conference on the use of Computers in Radiotherapy (ICCR). Amsterdam, Netherlands 2010.
 93. Miften M, Olch A, Mihailidis D, et al. Tolerance limits and methodologies for IMRT measurement-based verification QA: recommendations of AAPM Task Group No. 218. *Med Phys.* 2018;45(4):e53-e83.
 94. Nag M, Liu J, Liu L, et al. *Body Location Embedded 3D U-Net (BLE-U-Net) for Ovarian Cancer Ascites Segmentation on CT Scans.* SPIE; 2023.
 95. Liu L, Liu J, Nag MK, et al. Improved Multi-modal Patch Based Lymphoma Segmentation with Negative Sample Augmentation and Label Guidance on PET/CT Scans. Third International Workshop, MMMI 2022, Held in Conjunction with MICCAI 2022, Singapore, September 22, 2022, Proceedings Sep 2022; pp 121–129.

How to cite this article: Ding Y, Feng H, Yang Y, et al. Deep-learning based fast and accurate 3D CT deformable image registration in lung cancer. *Med. Phys.* 2023;1-17.
<https://doi.org/10.1002/mp.16548>