Inter-fraction deformable image registration using unsupervised deep learning for CBCT-guided abdominal radiotherapy

Huiqiao Xie, Emory University
Yang Lei, Emory University
Yabo Fu, Emory University
Tonghe Wang, Emory University
Justin Roper, Emory University
Jeffrey Bradley, Emory University
Pretesh Patel, Emory University
Tian Liu, Emory University
Xiaofeng Yang, Emory University

Journal Title: Physics in medicine and biology
Volume: Volume 68, Number 9
Publisher: IOP Publishing | 2023-04-13
Type of Work: Article | Final Publisher PDF
Publisher DOI: 10.1088/1361-6560/acc721
Permanent URL: https://pid.emory.edu/ark:/25593/w3zch

Final published version: http://dx.doi.org/10.1088/1361-6560/acc721

Copyright information:
© 2023 The Author(s). Published on behalf of Institute of Physics and Engineering in Medicine by IOP Publishing Ltd

This is an Open Access work distributed under the terms of the Creative Commons Attribution 4.0 International License (https://creativecommons.org/licenses/by/4.0/rdf).

Accessed June 7, 2024 8:12 AM EDT
Inter-fraction deformable image registration using unsupervised deep learning for CBCT-guided abdominal radiotherapy

Huiqiao Xie1, Yang Lei, Yabo Fu1,2, Tonghe Wang1,2, Justin Roper1, Jeffrey D Bradley1, Pretesh Patel1, Tian Liu1,3 and Xiaofeng Yang1,∗

1 Department of Radiation Oncology and Winship Cancer Institute, Emory University, Atlanta, GA, United States of America
2 Department of Medical Physics, Memorial Sloan Kettering Cancer Center, New York, NY, United States of America
3 Department of Radiation Oncology, Icahn School of Medicine at Mount Sinai, New York, NY, United States of America

∗ Author to whom any correspondence should be addressed.
E-mail: xiaofeng.yang@emory.edu

Keywords: deformable image registration, cbct, radiotherapy, deep learning

Abstract

Objective. CBCTs in image-guided radiotherapy provide crucial anatomy information for patient setup and plan evaluation. Longitudinal CBCT image registration could quantify the inter-fractional anatomic changes, e.g. tumor shrinkage, and daily OAR variation throughout the course of treatment. The purpose of this study is to propose an unsupervised deep learning-based CBCT-CBCT deformable image registration which enables quantitative anatomic variation analysis. Approach. The proposed deformable registration workflow consists of training and inference stages that share the same feed-forward path through a spatial transformation-based network (STN). The STN consists of a global generative adversarial network (GlobalGAN) and a local GAN (LocalGAN) to predict the coarse- and fine-scale motions, respectively. The network was trained by minimizing the image similarity loss and the deformable vector field (DVF) regularization loss without the supervision of ground truth DVFs. During the inference stage, patches of local DVF were predicted by the trained LocalGAN and fused to form a whole-image DVF. The local whole-image DVF was subsequently combined with the GlobalGAN generated DVF to obtain the final DVF. The proposed method was evaluated using 100 fractional CBCTs from 20 abdominal cancer patients in the experiments and 105 fractional CBCTs from a cohort of 21 different abdominal cancer patients in a holdout test. Main Results. Qualitatively, the registration results show good alignment between the deformed CBCT images and the target CBCT image. Quantitatively, the average target registration error calculated on the fiducial markers and manually identified landmarks was 1.91 ± 1.18 mm. The average mean absolute error, normalized cross correlation between the deformed CBCT and target CBCT were 33.42 ± 7.48 HU, 0.94 ± 0.04, respectively. Significance. In summary, an unsupervised deep learning-based CBCT-CBCT registration method is proposed and its feasibility and performance in fractionated image-guided radiotherapy is investigated. This promising registration method could provide fast and accurate longitudinal CBCT alignment to facilitate inter-fractional anatomic changes analysis and prediction.

1. Introduction

The treatment course of fractionated radiotherapy usually lasts weeks or months, during which the shape, position and size of tumor targets and organs at risk (OARs) may vary to different extents. It is essential to account for these changes for accurate dose delivery (Yan 2010, Thörnqvist et al 2016). Cone-beam computed tomography (CBCT) is commonly used in image-guided radiotherapy for patient setup and treatment evaluation prior to beam delivery. Compared to the treatment planning CT, CBCT images usually have poor image quality and soft tissue contrast due to scattering and beam hardening. Since the on-treatment CBCT scan
with gantry mounted kV imaging on Linacs usually takes longer time (around one minute) than the planning CT (several seconds). The CBCT images are also more susceptible to motion artifacts. As a result, CBCTs are primarily used to verify the alignment with planning CT. The ability of CBCT to track patients’ anatomic changes throughout the treatment course has been under-explored. CBCT could enable detailed inter-fractional anatomic changes evaluation on treatment day throughout the treatment course. As such, inter-fractional target and OAR variation could potentially be modeled to predict anatomic variation in future treatment or to guide treatment planning, such as margin definition, to provide sufficient target coverage and OAR sparing. It is important to provide a CBCT-CBCT image registration tool to quantify the inter-fractional anatomic changes to facilitate such modeling.

Conventional intensity-based deformable image registrations (DIRs), such as optical flow (Østergaard Noe et al 2008), demons (Wang et al 2005) and viscous fluid model (D’Agostino et al 2003), are iterative and generally very slow especially for large datasets. Over-smoothed deformation vector fields (DVFs) are usually produced because these methods apply spatial filters repeatedly throughout the iteration process (Yang et al 2010, Fu et al 2018). The large appearance variances and low image contrast of CBCT pose additional challenges for accurate registration. Landmarks identified either automatically or manually have been used to guide planning CT-CBCT and MRI-CBCT registration (Kearney et al 2014, Fu et al 2020b). However, the landmark identification process can be challenging and laborious in the presence of severe artifacts, which in turn degrade the landmark-guided DIR (Motegi et al 2019). Partly due to the poor CBCT image quality, very few papers have been published on CBCT-CBCT image registration (Østergaard Noe et al 2008, Nithiananthan et al 2009, Zachiu et al 2017, Jiang et al 2020). Noe et al (Østergaard Noe et al 2008) tried to accelerate an optical flow method (Cornelius and Kanade 1984) on a graphics programming unit (GPU) and the test on CBCT-CBCT registration achieved run time of 64 s for image size of 512 × 512 × 55 and target registration errors (TREs) of 1.8 ± 0.10 mm after rigid registration and 1.6 ± 0.8 mm after deformable registration. Nithiananthan et al (2009) studied the accuracy and convergence of multiscale Demons image registration and they achieved run time of 270 s (image size was not specified) and TRE of 1.6 ± 0.9 mm on the CBCT images of ten head and neck cancer patients. Zachiu et al (2017) implemented and evaluated an Evolution method (Denis de Senneville et al 2016), which estimates the deformation between two images by matching similar contrast patterns instead of pixel intensities, for both CT-CBCT and CBCT-CBCT image registration. They achieved run time of approximately 60 s for the registration of images of size 256 × 256 × 256. Jiang et al (2020) proposed a multi-scale deformable image registration (DIR) framework with unsupervised joint training of convolutional neural network (MJ-CNN) for 4D-CT inter-phase registration. It was shown that, though being trained on a 4D-CT dataset, the MJ-CNN framework also performed well on both CT-CBCT and CBCT-CBCT registration without re-training or fine-tuning with a run time of about 1.4 s for an image size of 256 × 256 × 96.

Recently, Deep learning (DL)-based medical image registration has become a hot research topic and achieved promising performances. Two thorough review papers on DL-based image registration were recently published by Fu et al (2020a) and Haskins et al (2020). Generally, DL-based image registration methods can be divided into three categories: deep iterative registration, supervised transformation prediction and unsupervised transformation prediction. The limitation of deep iterative registration is that they inherit the iterative nature of conventional DIR methods (Haskins et al 2019), which slows the registration process. Supervised transformation prediction methods utilize the ‘ground truth DVF’ which is usually obtained using other DIR methods or artificially generated and thereafter quality checked by experts to supervise the network training (Pei et al 2017). The quality control of the fidelity of the ‘ground truth DVF’ are subjective which may induce further inter-observer variability of the network performance. Unsupervised transformation prediction enables a large number of datasets to be used in training since no ‘ground truth DVF’ is needed. However, without ground truth transformations, it is difficult to define proper loss functions of the networks. A spatial transformer network (STN) (Jaderberg et al 2015) was proposed to generate the deformed image which enables image similarity loss calculation during the training process.

In this study, a novel unsupervised deep learning framework for DIR of inter-fraction CBCT images is proposed. Several strengths in the network design are considered in the proposed workflow: (I) This work is based on the STN which explicitly enables the loss function to be defined without any manually aligned or pre-registered image pairs (Jaderberg et al 2015, Lei et al 2020). Loss functions of image similarity and DVF regularizations are used in the training stage of the proposed workflow. (II) GAN architectures have been incorporated in the STN to improve the realism of the predicted DVFs. (III) A multi-scale framework was adopted to capture the coarse-scale and fine-scale motion.

2. Methods and materials

2.1. The proposed workflow of spatial transformation-based network

The schematic flowchart of the proposed method is outlined in figure 1. The training and inference stages follow the same feed-forward path through an STN, which consists of a global generative adversarial network.
GlobalGAN and a local GAN (LocalGAN). The GlobalGAN is trained with the whole volume of the moving and the target images to capture the global geometric deformation and to generate a global deforming vector field (DVF) which facilitates the coarse alignment. The global DVF and moving CBCT fraction are then fed into spatial transformation to generate the globally deformed fractional CBCT images. However, the global DVF may fail to provide accurate local image registration due to non-rigid geometry and anatomic movements. To improve the accuracy of local registration, a LocalGAN is designed to capture the local deformation on top of the globally deformed CBCT images to match the target image. In the training of the LocalGAN, three-dimensional (3D) image patches with a size of 64 $\times$ 64 $\times$ 64 are extracted from the globally deformed CBCT images and the target CBCT images with an overlap size of 32 $\times$ 32 $\times$ 48. An image similarity loss and a regularization loss, as well as an adversarial loss, are included in the loss function of the GlobalGAN and LocalGAN. During the training stage, the GlobalGAN and LocalGAN are trained without the supervision of ground truth DVFs.

In the inference stage, the global DVF and patches of local DVF are sequentially predicted by the GlobalGAN and LocalGAN. The patches of local DVF are then tiled and averaged to generate the whole-volume local DVF. The final DVF can be obtained by combining the global DVF and the whole-volume local DVF.

Our algorithm was implemented in Python 3.6.9 and Tensorflow 1.8 with Adam gradient descent optimizer and was trained and tested on a NVIDIA Tesla V100 GPU with 32 GB of memory. We also used several libraries and toolboxes such as numpy, scikit-image, pydicom, h5py, and scipy. The inference time for a new patient’s multi-CBCTs is about 5 min and depends on the number of fractions.

2.1.1. Architectures of the networks
The GlobalGAN and LocalGAN share similar GAN architectures with different learnable parameters. The structures of the generator and discriminator are shown in figure 2. Architecture details of the networks can be found in tables A.1 and A.2 of the appendix. In the generator network, image sizes of the input image pairs are reduced while being encoded through 11 convolutional layers. In order to up-sample the generated DVFs to matrix sizes the same as the input images, bilinear interpolation was applied. The discriminators in the two GANs are implemented as conventional fully convolution networks (FCN) (Lei et al 2019) for the necessary regularization to generate realistic DVFs. Since the discriminators are trained to distinguish the deformed images from the real CBCT images, they encourage the GlobalGAN and LocalGAN to predict realistic DVFs by penalizing unrealistic deformed images.

2.1.2. Attention gates
In order to force the model to focus on learning the motion information, self-attention network is integrated into GAN architectures for both GlobalGAN and LocalGAN. Self-attention network is constructed by integrating two attention gates into the generator between the convolution layers before and after the max pooling layers. With the attention gates, feature maps with different scales of the adjacent convolution layers are combined and operated right before concatenation. Attention gates have previously been explored in the context...
of semantic segmentation and were able to capture the most relevant semantic contextual information without using a very large receptive field (Romera-Paredes and Torr 2016, Oktay et al 2018).

2.1.3. Loss functions
The image similarity, regularization and adversarial losses are consisted in the loss function of the GlobalGAN and LocalGAN. The difference between the loss term used for GlobalGAN and LocalGAN is that the loss of GlobalGAN is calculated based on whole images and the loss of LocalGAN is calculated based on patches.

\[
G = \arg \min_G \{ \alpha \cdot (\text{SIM}(I_d, I_t)) + \beta \cdot \text{ADV}(I_d, I_t) + \gamma \cdot \text{R}(\text{DVF}) \},
\]

where DVF = G(I_m, I_t) represents the predicted DVF from a pair of moving I_m and target I_t images. The deformed image I_d can then be derived by applying DVF to the moving image I_m as \( I_d = \text{DVF} \odot I_m \) where \( \odot \) denotes the operation of applying DVF to the moving image.

\[
\text{SIM}(I_d, I_t) = [1 - \text{NCC(MIND}(I_d), \text{MIND}(I_t))] + \delta \cdot \text{GD(MIND}(I_d), \text{MIND}(I_t))
\]

where NCC(·) and GD(·) denote the normalized cross-correlation (NCC) loss and the gradient difference (GD) loss, respectively, between the deformed and target images. MIND(·) denotes the modality independent neighbourhood descriptor (MIND) (Heinrich et al 2012). Due to scattering and image artifacts, inter-fraction CBCT HU values are inconsistent, which may deteriorate the effectiveness of similarity metrics such as mean square error and NCC. The MIND descriptor is a modality independent neighbourhood descriptor with normalized image intensity which is used here to pre-process the images before similarity measurement using NCC and GD.

ADV(·) denotes the adversarial loss that is computed as the discriminator binary cross entropy loss of the deformed and target images. The purpose of the adversarial loss is to encourage the deformed image to approach realistic CBCT image by penalizing unreasonable DVFs and unrealistic deformed images.

R(DVF) denotes the regularization term

\[
\text{R}(\text{DVF}) = \mu_1 \nabla^2 \text{DVF}_1^2 + \mu_2 \nabla^2 \text{DVF}_2^2.
\]

Weighted first and second derivatives of the DVF are included in the regularization term to enforce general smoothness of the predicted DVF. Values of \( \mu_1 \) and \( \mu_2 \) are set as 1 and 0.5 in this study, respectively.

The hyperparameters of \( \alpha, \beta, \gamma \) and \( \delta \) are empirically set as 200, 1, 10 and 5, respectively, according to numerical experiments.

2.2. Datasets and experiments
100 fractional CBCT images of 20 abdominal (pancreas and liver) cancer patients who underwent radiotherapy were retrospectively investigated. The CBCT images have a resolution of 0.90 mm × 0.90 mm × 2.0 mm with size of 512 × 512 × 88 and are acquired by the on-board imaging system mounted on Varian TrueBeam linear accelerators (Varian Medical Systems, Inc., Palo Alto, CA). The built-in adult Pelvis protocol, with 125kVp, 60 mA, 20 ms per pulse and 1080 mAs per scan, was selected by the therapists before each fraction. These images were acquired during a five-fraction treatment course of each patient. Fiducial markers were implanted in the patients for tumor localization and external beam treatment planning.
The overall performance of the proposed method was investigated via a five-fold cross-validation. Specifically, the CBCT image data of the 20 patients was first randomly and equally separated into five groups, of which four groups were used for training; and the rest group was used for testing. The training and testing experiments were repeated five times by rotating each group as the testing group.

In a holdout test, 105 fractional CBCTs from a cohort of 21 different abdominal cancer patients were investigated to evaluate the proposed method. The trained STN with the original 100 fractional CBCT images of the 20 abdominal cancer patients was tested on the holdout dataset without re-training or parameter fine-tuning.

This research has been approved by IRB without informed notices.

2.3. Evaluations

Qualitative evaluations of the proposed method were performed by visually assessing the alignment between the target and deformed CBCT images. Both the fusion images and the absolute difference images between the target and deformed images were generated for the visual assessment. The absolute intensity difference profiles along a line in the anterior-posterior direction are also plotted to demonstrate the accuracy of image alignment. To demonstrate the efficacy of the integrated attention gates, DIR results with and without the attention gates were compared.

For quantitative evaluations, mean absolute errors (MAEs) and normalized cross correlations (NCCs) between the target and the deformed CBCT images were calculated. The TREs and dice similarity coefficients (DSCs) were also calculated. The Jacobian determinants of the DVFs (Brock et al. 2017) were calculated to assess the fidelity of the predicted DVFs. To further evaluate the DVF’s reasonability, the regularity of the registration field (DVF, \( \varphi \)) was also evaluated. To be specific, for each voxel \( i \), the Jacobian matrix \( \mathbf{J} (\varphi(i)) = \nabla \varphi(i) \), which captures the local properties of DVF \( \varphi \) around voxel \( i \), was calculated. Then the ratio (denoted by \( \%N \)) of all non-background voxels (within body contour) was calculated for which \( J(\varphi(i)) \leq 0 \), where the deformation is not diffeomorphic, to the total number of non-background voxels. As such, lower ratio indicates better DVF regularity.

The TRE was calculated as Euclidean distance between the landmark positions in the target and deformed CBCT images. The implanted fiducial markers with two or more additional landmarks identified by an experienced medical physicist are used for the TRE calculation. Four examples of the selected landmarks are shown in figure 3(a) as red ‘×’ marks, of which two were implanted fiducials and the other two were the tips of the spinous process and rib bone. These landmarks were selected due to the fact that physicians usually prescribe to ‘match to target’ or ‘match to bone’ for on-treatment patient setup. With the position of the 7th landmark in patient \( K \) being denoted as \( P^j_K \) in the moving fraction and the position of its corresponding landmark in the same patient denoted as \( \hat{P}^j_K \), the TRE was calculated as:

\[
\text{TRE}(i, K) = \| P^j_K - \hat{P}^j_K \|_2,
\]

where \( \| \cdot \|_2 \) stands for the L-2 norm of the matrix \( \cdot \).

The MAE for patient \( K \) was calculated as

\[
\text{MAE}(K) = \frac{1}{\| B_K \|_0} B_K \| (I_{K,d} - I_{K,t}) \|_1,
\]

where the target and deformed images of patient \( K \) are denoted as \( I_{K,t} \) and \( I_{K,d} \). \( B_K \) stands for the image mask of the bounding box of the patient body, which was determined by all the tissues/organs with HU values higher than -300 HU in this study.

The NCC for patient \( K \) was calculated as

\[
\text{NCC}(K) = \frac{\sum_{x,y,z} (B_K (I_{K,d}) - \bar{B}_K (I_{K,d})) \cdot (B_K (I_{K,t}) - \bar{B}_K (I_{K,t}))}{\sqrt{\sum_{x,y,z} (B_K (I_{K,d}) - \bar{B}_K (I_{K,d}))^2 \sum_{x,y,z} (B_K (I_{K,t}) - \bar{B}_K (I_{K,t}))^2}}^{1/2},
\]

where \( \sum_{x,y,z} \) stands for elemental wise summation of the 3D CBCT images; and \( B_K \) stands for the mean of the image \( B \). \( \bar{B}_K \) has the same meaning as in the calculation of MAE of equation (5).

The DSC for patient \( K \) was calculated as

\[
\text{DSC}(K) = \frac{2 \| M_{K,d} \cdot M_{K,t} \|_0}{\| M_{K,d} \|_0 + \| M_{K,t} \|_0},
\]

where \( M_{K,d} \) and \( M_{K,t} \) are the binary mask of the bony structures in the deformed and target images. \( M_{K,d} \cdot M_{K,t} \) stands for the element-wise multiplication of these masks. The masks of the bony structures for the DSC calculation in this study were determined by the tissues/organs with HU values higher than 300 HU. Calculation of the DSC on bony structures could add an evaluation of the image registration, i.e. residual alignment and absence of undesired bone deformation, in addition to the TREs, which are landmark points based.
3. Results

Fractional CBCT images of an abdominal cancer patient during the treatment course are shown in figure 3 to demonstrate the overall registration results of the proposed method. In figure 3, the first fraction is shown in the upper left corner as the target CBCT image; and the subsequent four fractions (the following four columns) that were registered with on-treatment manual rigid registration, the proposed method without attention gates in the generators of the GlobalGAN and LocalGAN and the proposed method are shown in the first to third rows, respectively. Three observations could be drawn from figure 3: (1) The stomach region appearances could be essentially different in each treatment fraction (pointed by orange arrows); (2) The overall body (anterior-posterior) location could be well resemble to the target image with the proposed method, but not the manual rigid registration or the proposed method without attention gates (regions pointed by yellow arrows); (3) The anatomical morphology of the internal organs, fiducial markers, and even the streak artifacts, are also changed/modified, which is believed to be induced by the LocalGAN. Overall, the proposed method has better global and local alignment due to the multi-scale registration scheme realized by the GlobalGAN and LocalGAN.

For better visibility of the registration quality, the fusion images of the registered fractions with the target fraction of the same patient shown in figure 3 at a different slice location are shown in figure 4. In the fusion images, the deformed images and the target image are show in the red and green channels, respectively; and they are labeled as the same subfigures as in figure 3 for consistence. Since the stomach region appearance was different from fraction to fraction, suboptimal image registration quality can be found (pointed by orange arrows). As shown by the yellow arrows in figure 4, with the proposed method, the liver area could be well registered to the target region, but not with the other two methods. Bony structures are also usually treated as indicators of image registration quality. For the ribs and the spine, (regions pointed by red arrows), the proposed method (c2–c5) outperformed the other two approaches (a2–a5, b2–b5).

The difference images between the registered images (a2–a5, b2–b5 and c2–c5) and the target images (a1) in figure 4 are shown in figure 5. The difference images are labeled as the same subfigures as in figure 4; and they have the same arrows copied from figure 4. The proposed method has shown better alignment both globally and locally, evidenced by the lower HU difference, than the other two methods. Despite of the challenges at the stomach region, the proposed method shows the lowest HU difference (orange arrows). Consistent with previous observations, the proposed method outperformed the other two methods at the liver and spine boundaries, as indicated by the yellow and red arrows.

Shown in figure 6 are the absolute HU difference profiles along the black dotted line in figure 5. The large HU difference anteriorly (voxel index <20) and posteriorly (voxel index >300) were caused by the body contour.
misalignment; and those near the spine ($240 < \text{voxel index} < 300$) were caused by the spine misalignment.

Compared with the other two methods, the proposed method has the best performance with lowest HU differences. The average TREs over all patients in the five-fold cross validation are listed in table 1. The average TREs is less than 2 mm for the proposed method, much better than the other compared methods (on-treatment manual rigid registration).

Figure 4. Fusion images of the target fraction and the registered fractions of the patient shown in figure 3 at a different slice location. The fusion images of first (target) fraction and the second to fifth fractions with on-treatment manual rigid registration are shown in the first row (a2–a5), respectively. The fusion images of the target fraction and those registered with the proposed method without attention gates are shown in the second row (b2–b5). The fusion images of the target fraction and those registered with the proposed method are accordingly shown in the third row (c2–c5).

Figure 5. Difference images of the registered second to fifth fractions versus the target fraction of the patient shown in figure 4. The difference images of the second to fifth fractions with on-treatment manual rigid registration are shown in the first row (a2–a5), respectively. The difference images of the registered fractional images with the proposed method without attention gates are shown in the second row (b2–b5). The difference images of those registered with the proposed method are accordingly shown in the third row (c2–c5).
rigid registration, the proposed method without attention gate, the proposed method without adversarial loss in equation (1), the proposed method without LocalGAN (GlobalGAN only) and the VoxelMorph method (Balakrishnan et al 2019)). The smaller TRE standard deviations of the proposed method also demonstrate better registration robustness. Two sample t-tests on the TRE results of the investigated methods showed that the TRE improvement of the proposed method over the others is statistically significant. The TRE improvement of the proposed method without attention gates over the on-treatment manual rigid registration was not statistically significant, showing the efficacy of the attention gates.

The average MAEs between the deformed CBCT images and the target images are listed in table 2. Same conclusion as the TRE analysis can be drawn, which strengthened our argument that the proposed method with attention gates has the best performance out of the three methods.

As two important metrics indicating the image registration quality, the NCCs and DSCs were calculated and shown in tables 3 and 4, respectively. From these two tables, same conclusion that the proposed method substantially outperformed the other two methods with statistical significance can be drawn.

Table 5 shows the average TREs, MAEs, NCCs and DSCs for all the patients in the holdout dataset. The uniform improvement of the proposed method over the other methods across all four metrics demonstrated that the proposed method has great robustness and generalizability on the holdout datasets.

The Jacobian determinant index of DVFs (Brock et al 2017) derived by the proposed method and comparing methods is shown in table 6. The superior performance of the proposed method over the ablation studies and comparing method (VoxelMorph) was demonstrated by the relatively smaller Jacobian determinant value, indicating improved topology preservation and DVF regularization. It is shown that the proposed method with only GlobalGAN has the largest Jacobian determinant index and the proposed method has the lowest Jacobian determinant index. According to the %N numbers, it shows that with the adversarial loss, the proposed method improves DVF regularity as compared to without using the adversarial loss. Additionally, compared to VoxelMorph, the proposed method can reach a lower percentage of non-positive DVF.

4. Discussions

Over the course of radiotherapy treatment, a method to provide fast and accurate inter-fraction CBCT image registration is essential for evaluation of the geometric and anatomic changes. With the proposed registration tool, quantitative anatomic changes could be calculated for inter-fractional variation modeling and prediction. It could potentially inform the physician in future treatment planning such as targets margin definition and image guidance usage frequency, tradeoff between target coverage and OAR sparing. The proposed inter-fraction CBCT image registration could also enable many applications such as image segmentation (Wang et al 2020), motion estimation (Fu et al 2018, Kai et al 2020), image fusion (Fu et al 2021) and treatment response evaluations (Ou et al 2015, Posiewnik and Piotrowski 2019). Deep learning-based DIR is promising for the online DIR task of large volume CBCT images in radiotherapy. In this work, an unsupervised deep learning based inter-fraction CBCT registration method, which takes less than 3 s to perform a CBCT-CBCT registration, is proposed and its feasibility and performance are investigated through qualitative and quantitative evaluations.
Table 1. Target registration errors (TREs) of the registered second to fifth fractions, as well as overall TRE regardless of fraction, versus the target fraction. The TREs are calculated over all of the involved patients in the experiment. (Unit: mm).

| Fraction | Manual rigid | Proposed w/o AG | Proposed w/o adv. loss | Proposed GlobalGAN only | VoxelMorph | Proposed 4 | p value (1 versus 2) | p value (1 versus 4) (2 versus 4) (3 versus 4) |
|----------|--------------|-----------------|------------------------|-------------------------|------------|-----------|----------------------|-----------------------------------------------|
| 2        | 4.85 ± 2.73  | 4.46 ± 2.34     | 4.12 ± 2.18            | 5.58 ± 2.89             | 4.11 ± 2.19| 1.92 ± 0.97 | 0.612                | <0.01                                        |
| 3        | 4.24 ± 2.69  | 3.66 ± 2.26     | 4.36 ± 2.37            | 5.65 ± 1.54             | 4.02 ± 2.12| 1.74 ± 0.95 | 0.409                | <0.01                                        |
| 4        | 6.28 ± 3.72  | 5.35 ± 3.66     | 3.78 ± 2.17            | 5.18 ± 2.16             | 3.79 ± 1.98| 1.99 ± 1.14 | 0.379                | <0.01                                        |
| 5        | 4.88 ± 2.70  | 5.07 ± 3.35     | 4.50 ± 2.26            | 5.68 ± 1.82             | 4.49 ± 1.63| 1.97 ± 1.59 | 0.836                | <0.01                                        |
| Overall  | 5.07 ± 3.05  | 4.64 ± 3.01     | 4.24 ± 2.23            | 5.32 ± 2.07             | 4.15 ± 2.15| 1.91 ± 1.18 | 0.320                | <0.01                                        |
Table 2. Mean absolute errors (MAEs) of the registered second to fifth fraction images, as well as overall MAE regardless of fraction, versus the target fraction image. The MAEs are calculated over all of the involved patients in the experiment. (Unit: HU).

| Fraction | Manual rigid | Proposed w/o AG | Proposed$^\dagger$ | $p$ value (1 versus 2) | $p$ value (1 versus 3) | $p$ value (2 versus 3) |
|----------|--------------|-----------------|--------------------|------------------------|-----------------------|-----------------------|
| 2        | 48.25 ± 13.80 | 44.42 ± 10.04   | 34.44 ± 7.73       | 0.359                  | 0.013                 | 0.007                 |
| 3        | 45.29 ± 10.74 | 41.40 ± 10.07   | 33.35 ± 8.63       | 0.331                  | 0.016                 | 0.032                 |
| 4        | 44.45 ± 9.83  | 42.74 ± 9.07    | 33.09 ± 7.49       | 0.620                  | 0.009                 | 0.005                 |
| 5        | 43.15 ± 9.40  | 42.64 ± 8.95    | 32.81 ± 7.08       | 0.881                  | 0.012                 | 0.004                 |
| Overall  | 45.28 ± 10.81 | 42.80 ± 9.47    | 33.42 ± 7.48       | 0.179                  | $<0.001$              | $<0.001$              |

Table 3. Normalized cross correlations (NCCs) of the registered second to fifth fractions, as well as overall NCC regardless of fraction, versus the target fraction. The NCCs are calculated over all of the involved patients in the experiment. Only tissues/organs with voxel intensity higher than −300 HU are considered.

| Fraction | Manual rigid | Proposed w/o AG | Proposed$^\dagger$ | $p$ value (1 versus 2) | $p$ value (1 versus 3) | $p$ value (2 versus 3) |
|----------|--------------|-----------------|--------------------|------------------------|-----------------------|-----------------------|
| 2        | 0.88 ± 0.06  | 0.89 ± 0.05     | 0.93 ± 0.04        | 0.345                  | $<0.001$              | $<0.001$              |
| 3        | 0.90 ± 0.06  | 0.90 ± 0.06     | 0.94 ± 0.04        | 0.815                  | $<0.001$              | $<0.001$              |
| 4        | 0.89 ± 0.04  | 0.89 ± 0.05     | 0.94 ± 0.03        | 0.963                  | $<0.001$              | $<0.001$              |
| 5        | 0.90 ± 0.05  | 0.89 ± 0.04     | 0.94 ± 0.04        | 0.230                  | $<0.001$              | $<0.001$              |
| Overall  | 0.89 ± 0.05  | 0.89 ± 0.05     | 0.94 ± 0.04        | 0.880                  | $<0.001$              | $<0.001$              |

Table 4. Dice similarity coefficients (DSCs) of the registered second to fifth fractions, as well as overall DSC regardless of fraction, versus the target fraction. The DSCs are calculated over all of the involved patients in the experiment. Only bony tissues/organs with voxel intensity higher than 300 HU are considered.

| Fraction | Manual rigid | Proposed w/o AG | Proposed$^\dagger$ | $p$ value (1 versus 2) | $p$ value (1 versus 3) | $p$ value (2 versus 3) |
|----------|--------------|-----------------|--------------------|------------------------|-----------------------|-----------------------|
| 2        | 0.34 ± 0.12  | 0.36 ± 0.10     | 0.51 ± 0.10        | 0.536                  | 0.004                 | $<0.001$              |
| 3        | 0.36 ± 0.12  | 0.37 ± 0.11     | 0.52 ± 0.11        | 0.716                  | 0.005                 | $<0.001$              |
| 4        | 0.35 ± 0.08  | 0.36 ± 0.08     | 0.52 ± 0.07        | 0.872                  | $<0.001$              | $<0.001$              |
| 5        | 0.36 ± 0.11  | 0.37 ± 0.10     | 0.52 ± 0.09        | 0.878                  | 0.003                 | $<0.001$              |
| Overall  | 0.35 ± 0.10  | 0.37 ± 0.10     | 0.52 ± 0.09        | 0.495                  | $<0.001$              | $<0.001$              |

The proposed method can also perform well on the holdout dataset without re-training. The major contributions of the proposed workflow can be summarized as:

(a) A multi-scale unsupervised deep learning-based DIR method is proposed for inter-fraction CBCT DIR in image-guided radiotherapy. The unsupervised training of the proposed STN-based network overcomes the challenge of collecting large amount of ground truth datasets via either manually aligning, which is labor-intensive, or artificially synthesizing, which is error prone. The integration of GlobalGAN and LocalGAN networks captures the image misalignment in a multi-scale manner; and its effects for the DIR can be observed in figures 3–6.

(b) Adversarial network is integrated into the proposed framework to enforce additional DVF regularization by penalizing unrealistic deformed images. Since the networks are designed to be trained in an unsupervised manner, DVF regularization is necessary to generate realistic results. Smoothness constraint has been commonly used in the literature for DVF regularization (Fu et al 2018). However, the smoothness constraint alone is insufficient for realistic DVF prediction especially when the network is trained in a completely unsupervised manner. Therefore, for additional DVF regularization, a discriminator is proposed to be integrated into STN. Since the purpose of the discriminator is to better differentiate deformed images from target images, such unrealistic deformed images are penalized. Realistic DVFs are then encouraged to be predicted by the generator. The speed of the inference stage will not be affected as the discriminator is only used in the training stage.

To generate DVFs with the same matrix sizes as the input images, since image sizes of the input image pairs are reduced while being encoded through 11 convolutional layers in the generator network, bilinear interpolation was used to up-sample the DVFs. As an alternative, transpose-convolution layers with trainable parameters can be used to up-sample the DVFs (Dumoulin and Visin 2016). However, we have found that bilinear interpolation, which does not contain trainable parameters, performs much better than the transpose-convolution layers in predicting accurate DVFs. The reason of this might be that bilinear interpolation tends to
Table 5. The overall TREs, MAEs, NCCs and DSCs of the registered CBCTs, regardless of fraction, versus the target CBCTs among all the patients in the holdout dataset. The methods for metric calculations are consistent with those in tables 1–5.

| Overall metrics | Manual rigid$^1$ | Proposed w/o AG$^2$ | Proposed w/o Adv. loss | Proposed GlobalGAN only | VoxelMorph$^3$ | Proposed$^4$ | $p$ value (1 versus 2) | $p$ value (1 versus 4) (2 versus 4) (3 versus 4) |
|-----------------|-----------------|---------------------|------------------------|-------------------------|----------------|---------------|-------------------|----------------------------------|
| TRE (mm)        | 5.12 ± 2.82     | 4.96 ± 2.40         | 4.39 ± 2.17            | 5.08 ± 2.58             | 4.71 ± 2.41   | 2.34 ± 1.74   | 0.229             | <0.01                            |
| MAE (HU)        | 50.74 ± 9.34    | 49.34 ± 9.15        | 43.70 ± 12.04          | 44.63 ± 12.08           | 42.14 ± 12.20 | 38.83 ± 7.88  | 0.499             | <0.05                            |
| NCC             | 0.89 ± 0.04     | 0.90 ± 0.04         | 0.90 ± 0.02            | 0.91 ± 0.03             | 0.89 ± 0.05   | 0.92 ± 0.02   | 0.037             | <0.01                            |
| DSC             | 0.41 ± 0.10     | 0.43 ± 0.09         | 0.41 ± 0.07            | 0.39 ± 0.09             | 0.44 ± 0.10   | 0.51 ± 0.04   | 0.341             | <0.01                            |
Table 6. The Jacobian determinant index of DVF derived by the proposed method and comparing methods.

|                                | Proposed w/o AG | Proposed w/o adv. loss | Proposed GlobalGAN only | VoxelMorph | Proposed  
|--------------------------------|----------------|------------------------|-------------------------|------------|---------
| Overall Jacobian determinant   | 0.007 ± 0.01   | 0.006 ± 0.008          | 0.014 ± 0.014          | 0.006 ± 0.007 | 0.005 ± 0.005  
| Non-positive Jacobian (%N)     | 0.311 ± 0.196  | 0.436 ± 0.170          | 0.279 ± 0.184          | 0.374 ± 0.159 | 0.298 ± 0.192  

generate smooth DVFs which are desired in medical image registration. On the other hand, the transpose-
convolution layers often generate unrealistic DVFs even with heavily-weighted DVF smoothness regularization term (Sokooti et al 2017, Lei et al 2020).

No prepossessing had been applied on the fractional CBCT images to improve the image quality before deformable image registration using the proposed method. Therefore, suboptimal image quality of the CBCT images, such as inter-fraction variations in the altered HU values and streak artifacts, could impact the accuracy of image registration. Deep learning-based image synthetic approaches have been investigated and shown promising result in improving the image quality (Wang et al 2021). Then, incorporating of these approaches with the proposed method might improve the image registration result. Further investigations on this topic are needed in future works.

One limitation about this work is that no landmark supervision was utilized in the training of the proposed network. Better deformable image registration results could be achieved if landmarks being selected according to the patient anatomy in CBCTs during the treatment course and being fed into the network to supervise the DVF generation. We anticipate to perform further investigation on this topic under collaboration with our physician group in the future. Another limitation of the proposed workflow is that no rigid registration preprocessing included. Since the network is designed only for generation of deformable image registration DVFs, unpredictable results could be generated if long coordinates shifts exist between the deformed and target images. This problem could be solved by incorporation of a rigid registration step before application of the proposed method.

It was observed that the inter-fraction shape and position of the organs in some abdominal patients may vary significantly due to gas fillings, bowel movements and/or respiratory motions. This is an extremely challenging situation for both unsupervised and supervised deep learning image registration, because that significant shape and position variations may become beyond model ability of the proposed method and may also impossible to manual labelling of the ground truth. The proposed method may fail to accurately register the inter-fraction CBCTs of these patients. Future researches are necessary for this situation.

Concerns still exist about the deep leaning-based methods for medical image registration, such as the topology preservation of the derived DVFs. To address this problem, inverse/cycle-consistent networks have been proposed to implicitly regularize the results of image registration (Zhang 2018, Kim et al 2021). It is expected that the proposed method could benefit from the same strategy.

In this study, only a few qualitative (visual inspections on the deformed, fusion, difference images and intensity profiles) and quantitative (TRE, MAE, NCC and DSC) evaluations were performed. These metrics are not directly related with clinical outcomes. It is anticipated that more clinical investigations, i.e. dose volume histograms, patient follow-ups, etc, are needed in order to concluded whether the proposed method could be applied and effective in the image-guided radiotherapy.

5. Conclusion

An unsupervised deep learning-based CBCT-CBCT registration method is developed and its feasibility and performance are investigated. The proposed method is able to accurately register images between the moving and target CBCT fractions within three seconds in a single forward network prediction, as such it is expected to be promising as a fast and straightforward image registration tool for motion management and treatment planning in image-guided radiotherapy.

Acknowledgments

This research is supported in part by the National Institutes of Health under Award Number R01CA215718, R56EB033332 and R01EB032680.

Data availability statement

The data cannot be made publicly available upon publication due to legal restrictions preventing unrestricted public distribution. The data that support the findings of this study are available upon reasonable request from the authors.

Conflicts of interests

The authors have no conflict of interests to declare.
Appendix. Architecture details of the generator and discriminator networks

### Table A.1. Architecture of the generator used in 2.

| Backbone   | Filter shape & stride size | # Channels | Padding |
|------------|----------------------------|------------|---------|
| Input      | N/A                        | 2          | N/A     |
| Conv_0     | $3 \times 3 \times 3 \& 1 \times 1 \times 1$ | 16         | SAME    |
| Conv_1     | $3 \times 3 \times 3 \& 1 \times 1 \times 1$ | 16         | SAME    |
| Conv_2     | $3 \times 3 \times 3 \& 1 \times 1 \times 1$ | 16         | SAME    |
| Concat.    | N/A                        | 48         | N/A     |
| Conv_3     | $3 \times 3 \times 3 \& 1 \times 1 \times 1$ | 48         | SAME    |
| Attention Gate_0 | $3 \times 3 \times 3 \& 1 \times 1 \times 1$ | 48         | SAME    |
| Conv_4     | $3 \times 3 \times 3 \& 1 \times 1 \times 1$ | 48         | SAME    |
| Max_pool_0 | $2 \times 2 \times 2$     | 48         | N/A     |
| Conv_5     | $3 \times 3 \times 3 \& 1 \times 1 \times 1$ | 96         | SAME    |
| Conv_6     | $3 \times 3 \times 3 \& 1 \times 1 \times 1$ | 96         | SAME    |
| Concat.    | N/A                        | 192        | N/A     |
| Conv_7     | $3 \times 3 \times 3 \& 1 \times 1 \times 1$ | 192        | SAME    |
| Max_pool_1 | $2 \times 2 \times 2$     | 192        | N/A     |
| Conv_8     | $3 \times 3 \times 3 \& 1 \times 1 \times 1$ | 384        | SAME    |
| Conv_9     | $3 \times 3 \times 3 \& 1 \times 1 \times 1$ | 384        | SAME    |
| Conv_10    | $3 \times 3 \times 3 \& 1 \times 1 \times 1$ | 384        | SAME    |
| Conv_final | $3 \times 3 \times 3 \& 1 \times 1 \times 1$ | 3          | SAME    |

### Table A.2. Architecture of discriminator used in figure 2.

| Backbone   | Filter shape & stride size | # Channels | Padding |
|------------|----------------------------|------------|---------|
| Input      | N/A                        | 1          | N/A     |
| Conv_0     | $9 \times 9 \times 9 \& 2 \times 2 \times 2$ | 8          | VALID   |
| Max_pool_0 | $2 \times 2 \times 2$     | 8          | N/A     |
| Conv_1     | $7 \times 7 \times 7 \& 2 \times 2 \times 2$ | 16         | VALID   |
| Max_pool_1 | $2 \times 2 \times 2$     | 16         | N/A     |
| Conv_2     | $5 \times 5 \times 5 \& 2 \times 2 \times 2$ | 32         | VALID   |
| Max_pool_2 | $2 \times 2 \times 2$     | 32         | N/A     |
| Conv_final | $3 \times 3 \times 3 \& 1 \times 1 \times 1$ | 2          | VALID   |

ORCID iDs

Tonghe Wang @ https://orcid.org/0000-0001-9021-1204
Xiaofeng Yang @ https://orcid.org/0000-0001-9023-5855

References

Balakrishnan G, Zhao A, Sabuncu M R, Guttag J and Dalca A V 2019 Voxelmorph: a learning framework for deformable medical image registration IEEE Trans. Med. Imaging 38 1788–800

Brock K K, Mutic S, McNutt T R, Li H and Kessler M L 2017 Use of image registration and fusion algorithms and techniques in radiotherapy: report of the AAPM radiation therapy committee task group no. 132 Med. Phys. 44 e43–76

Cornelius N and Kanade T 1984 Adapting optical-flow to measure object motion in reflectance and x-ray image sequences (abstract only) SIGGRAPH Comput. Graph. 18 24–5

D’Agostino E, Maes F, Vandermeulen D and Suetens P 2003 A viscous fluid model for multimodal non-rigid image registration using mutual information Med. Image Anal. 7 565–73

Denis de Senneville B, Zachius C, Ries M and Moonen C 2016 EVolution: an edge-based variational method for non-rigid multi-modal image registration Phys. Med. Biol. 61 7377–96

Dumoulin V and Visin F 2016 A guide to convolution arithmetic for deep learning. arXiv:1603.07285

Fu Y, Lei Y, Wang T, Curran W J, Liu T and Yang X 2020a Deep learning in medical image registration: a review Phys. Med. Biol. 65 20TR01

Fu Y, Liu S, Li H H, Li H and Yang D 2018 An adaptive motion regularization technique to support sliding motion in deformable image registration Med. Phys. 45 735–47

Fu Y, Wang T, Lei Y, Patel P, Jani A B, Curran W J, Liu T and Yang X 2020b Deformable MR-CBCT prostate registration using biomechanically constrained deep learning networks Med. Phys. 48 253–63
Fu Y, Wang T, Lei Y, Patel P, Jani A B, Curran W J, Liu T and Yang X 2021 Deformable MR-CBCT prostate registration using biomechanically constrained deep learning networks Med. Phys. 48 253–63
Haskins G, Kruecker J, Kruger U, Xu S, Pinto P A, Wood B J and Yan P 2019 Learning deep similarity metric for 3D MR–TRUS image registration Int. J. Comput. Assist. Radiol. Surg. 14 417–25
Haskins G, Kruger U and Yan P 2020 Deep learning in medical image registration: a survey Mach. Vis. Appl. 31 8
Heinrich M P, Jenkinson M, Bhushan M, Matin T, Gleggson F V, Brady S M and Schnabel J A 2012 MIND: Modality independent neighbourhood descriptor for multi-modal deformable registration Med. Image Anal. 16 1423–35
Jaderberg M, Simonyan K, Zisserman A and Kavukcuoglu K 2015 Spatial transformer networks arXiv:1506.02055
Jiang Z, Yin F F, Ge Y and Ren L 2020 A multi-scale framework with unsupervised joint training of convolutional neural networks for pulmonary deformable image registration Phys. Med. Biol. 65 015011
Kai Y, Arimura H, Niiomiyama K, Saito T, Shimohigashi Y, Kuraoka A, Maruyama M, Toya R and Oya N 2020 Semi-automated prediction approach of target shifts using machine learning with anatomical features between planning and pretreatment CT images in prostate radiotherapy J. Radiat. Res. 61 285–97
Kearney V, Chen S, Gu X, Chiu T, Liu H, Jiang L, Wang J, Yordy J, Nedzi L and Mao W 2014 Automated landmark-guided deformable image registration Phys. Med. Biol. 60 101–16
Kim B, Kim D H, Park S H, Kim J, Lee J G and Ye C 2021 CycleMorph: cycle consistent unsupervised deformable image registration Med. Image Anal. 71 102936
Lei Y et al 2019 Year MRI-based synthetic CT generation using deep convolutional neural network. Published SPIE Medical Imaging 10949 (San Diego, CA) 109492T
Lei Y, Fu Y, Wang T, Liu Y, Patel P, Curran W J, Liu T and Yang X 2020 4D–CT deformable image registration using multiscale unsupervised deep learning Phys. Med. Biol. 65 085003
Motegi K, Tachibana H, Motegi A, Hotta K, Baba H and Akimoto T 2019 Usefulness of hybrid deformable image registration algorithms in prostate radiation therapy J. Appl. Clin. Med. Phys. 20 229–36
Nithiananthan S, Brock K K, Daly M J, Chan H, Irish J C and Siewerdsen J H 2009 Demons deformable registration for CBCT-guided procedures in the head and neck: Convergence and accuracy Med. Phys. 36 4755–64
Oktay O et al 2018 Attention U-net: learning where to look for the pancreas arXiv:1804.03999
Ou Y et al 2015 Deformable registration for quantifying longitudinal tumor changes during neoadjuvant chemotherapy Magn. Reson. Med. 73 2343–56
Østergaard Noe K, De Seneville B D, Elstrom UV, Tanderup K and Serensen T S 2008 Acceleration and validation of optical flow based deformable registration for image-guided radiotherapy Acta Oncol. 47 1286–93
Pei Y, Zhang Y, Qin H, Ma G, Guo Y, Xu T and Zha H 2017 Year non-rigid craniofacial 2D–3D registration using CNN-Based regression, published (Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support, vol. Series) (Cham: Springer International Publishing) pp 117–25
Posiewnik M and Piotrowski T 2019 A review of cone-beam CT applications for adaptive radiotherapy of prostate cancer Phys. Med. 59 13–21
Romera-Paredes B and Torr PH S 2016 Year recurrent instance segmentation Published Computer Vision—ECCV 2016 ed B Leibe et al (Cham: Springer International Publishing) vol. Series pp 312–29
Sokooti H, de Vos B, Berendtse F, Lelieveldt B P F, Igum I and Staring M 2017 Year nonrigid image registration using multi-scale 3D convolutional neural networks. Published (Medical Image Computing and Computer Assisted Intervention — MICCAI 2017, vol. Series) (Cham: Springer International Publishing) pp 232–9
Thorsyöst S, Hysing L B, Tuomikoski L, Vestergaard A, Tanderup K, Muren L P and Heijmen B J 2016 Adaptive radiotherapy strategies for pelvic tumors - a systematic review of clinical implementations Acta Oncol. 55 943–58
Wang H, Dong L, O’Daniel J, Mohan R, Garden A S, Ang K K, Kuban D A, Bonnen M, Chang J Y and Cheung R 2005 Validation of an accelerated ‘demons’ algorithm for deformable image registration in radiation therapy Phys. Med. Biol. 50 2887–905
Wang T, Lei Y, Fu Y, Wynne J, Curran W J, Liu T and Yang X 2021 A review on medical imaging synthesis using deep learning and its clinical applications J. Appl. Clin. Med. Phys. 22 11–36
Wang Y, Zhong Z and Hua J 2020 Deeporgannet: on-the-fly reconstruction and visualization of 3D / 4D lung models from single-view projections by deep deformation network IEEE Trans. Visual Comput. Graphics 26 960–70
Yan D 2010 Adaptive radiotherapy: merging principle into clinical practice Semin. Radiat. Oncol. 20 79–83
Yang D, Goddu SM, Lu W, Pechenaya OL, Wu Y, Deasy JO, El Naqa I and Low DA 2010 Technical note: deformable image registration on partially matched images for radiotherapy applications Med. Phys. 37 141–5
Zachiu C, de Seneville B D, Tijsen R H N, Kotte A N T J, Houweling A C, Kerkmeijer L G W, Lagendijk J J W, Moonen C T W and Ries M 2017 Non-rigid CT/CBCT to CBCT registration for online external beam radiotherapy guidance Phys. Med. Biol. 63 015027
Zhang J 2018 Inverse-consistent deep networks for unsupervised deformable image registration arXiv:1809.03443