CNN-based real-time 2D-3D deformable registration from a single X-ray projection

François Lecomte, Jean-Louis Dillenseger, Stéphane Cotin

CNN-based real-time 2D-3D deformable registration from a single X-ray projection

F. Lecomte[1,2*], J.-L. Dillenseger[2,3] and S. Cotin[1,2]

[1*]Inria, 1 Place de l'Hopital, 67000, Strasbourg, France.
[2]ICube, CNRS, 1 Place de l'Hopital, 67000, Strasbourg, France.
[3]Rennes University, 3 rue du Clos Courtel, Rennes, France.

*Corresponding author: francois.lecomte@inria.fr;
Contributing authors: jean-louis.dillenseger@univ-rennes1.fr;
stephane.cotin@inria.fr;

Abstract

Purpose: The purpose of this paper is to present a method for real-time 2D-3D non-rigid registration using a single fluoroscopic image. Such a method can find applications in surgery, interventional radiology and radiotherapy. By estimating a three-dimensional displacement field from a 2D X-ray image, anatomical structures segmented in the preoperative scan can be projected onto the 2D image, thus providing a mixed reality view. **Methods:** A dataset composed of displacement fields and 2D projections of the anatomy is generated from the preoperative scan. From this dataset, a neural network is trained to recover the unknown 3D displacement field from a single projection image. **Results:** Our method is validated on lung 4D CT data at different stages of the lung deformation. The training is performed on a 3D CT using random (non domain-specific) diffeomorphic deformations, to which perturbations mimicking the pose uncertainty are added. The model achieves a mean TRE over a series of landmarks ranging from 2.3 to 5.5 mm depending on the amplitude of deformation. **Conclusion:** In this paper, a CNN-based method for real-time 2D-3D non-rigid registration is presented. This method is able to cope with pose estimation uncertainties, making it applicable to actual clinical scenarios, such as lung surgery, where the C-arm pose is planned before the intervention.

Keywords: 2D-3D registration, deformation, deep learning, real-time, fluoroscopy, diffeomorphism

1 Introduction

Laparoscopic surgery, interventional radiology and radiotherapy are among the most successful options for cancer therapy. These image-guided interventions are often totally or partially performed under fluoroscopic guidance, as this is for instance the case for lung surgery. On such moving and soft organs, the complexity of the intervention is obviously increased since the tumor position from the preoperative image becomes unknown or largely uncertain at the time of the procedure. The ability to manage organ motion (sometimes in real-time) is essential in the treatment outcome [1]. Therefore, estimating the new tumor position, in 3D and in real-time, from an intraoperative fluoroscopic image, can be a key improvement of these therapies. This is however very challenging as the combination of non-rigid deformation and reduced dimension of the intra-operative image make this problem ill-posed.

Several surveys cover the associated challenges and possible approaches to such 2D-3D registration problems [2, 3], as they can be handled in many different manners, given the application, imaging modality, parameter space, or optimization process. These registration methods can be divided in two groups: rigid and non-rigid. Rigid methods only compute a translation and a rotation while non-rigid methods are designed to estimate more complex displacement fields. Rigid registration methods are often needed to compensate for the unknown pose of the C-arm (and therefore the unknown projection from the CT to the fluoroscopic image) and can sometimes be sufficient when small deformations take place.

In image-guided radiotherapy techniques, a study by Kilburn *et al.* [4] showed that correcting for the intra-operative deformation increased the 2-year survival rate from 64 However, this procedure is invasive, increases the risks of complications to the patient and procedure time.

In the context of image-guided surgery, a study by Suzuki *et al.* [7] on Video-Assisted Thoracoscopic Surgery (VATs) found that failing to localize the tumor resulted in surgical conversion to open thoracotomy in 46

The Bayesian approach presented in [9] tackles the problem of 3D markerless tumor localization in radiotherapy. A pre-operative 4D CBCT is used to build a patient-specific respiratory model. During treatment, fluoroscopic images are acquired at frame rate of 5 images/s. A template matching algorithm is combined with a Kalman filter to predict the tumor position from the fluoroscopic image. Among the 13 cases in the study, the mean 3D error ranges from 1.6 to 2.9 mm. While these results are promising, the relatively high computation time and the need for a preoperative 4D CBCT reduces the clinical usability of this method.

The method in [10] uses a U-Net architecture combined with intensity-based registration and patient-specific biomechanical modeling to accurately localize liver tumors. The number of projections required, 20, requires specialized intraoperative imaging equipment, thus limiting the clinical applicability of the method.

Hirai *et al.* developed a Neural Network based markerless tumor localization method from fluoroscopic images [11]. The network is trained on Digitally Reconstructed Radiographs (DRRs) generated from a 4D-CT augmented by small rigid transformations (< 2 mm and < 1 deg). The authors report a 3D tumor localization accuracy of 1.6 mm, but the method requires the acquisition of a pair of fluoroscopic images.

The strategy proposed in [12] consists in applying deformations to the preoperative CT data paired with DRRs to form a training and testing dataset for a deep learning algorithm. Displacement fields are obtained from a 4D-CT by selecting a rest CT phase and registering the other phases to the rest phase.

A PCA is computed from the displacements fields and sampled to generate the patient-specific deformation dataset. A neural network is then trained to estimate the PCA components from the DRR input.

The authors only evaluated their method on data from the PCA that was used to train the network. No target registration error (TRE) is provided.

Most of the non-rigid registration methods described above assume a perfectly known intra-operative patient pose, which is needed to train the neural network. Many require a 4D-CT as preoperative image to generate patient-specific or problem-specific training data. Several methods also require multiple fluoroscopic images as input. In addition, existing methods are often limited in estimating a transformation from the 3D preoperative image space to the 2D fluoroscopic image. This is typically the case when a rigid transformation is estimated. However, this might prove insufficient when 3D tracking is required.

To address these limitations, we propose an approach that is robust against variation in poses, does not require a time series of preoperative images, and works with a single fluoroscopic image, in real-time. As a result our method is generic enough to address a variety of clinical applications where large, non-rigid deformations can occur.

To circumvent the need for a preoperative 4D CT, we implement a domain randomization solution to generate displacement fields using only the routinely acquired preoperative 3D CT. For this, we rely on the large deformation diffeomorphic metric mapping (LDDMM) framework (see section 2.2) to enforce smoothness and invertibility of the displacement fields. We then introduce a small pose variation of the C-arm to account for the discrepancy between the planned and actual intra-operative pose. We generate DRRs from the deformed CT volumes.

Our fully convolutional neural network then learns the mapping between the DRR and the 3D displacement field representing the rigid + non-rigid transform (see section 2.1). The resulting 3D displacement field can be applied to the preoperative CT, or a segmentation extracted from it, to provide real-time tracking or visualization of internal structures (see section 3).

2 Method

In order to be clinically relevant, our registration framework is based upon the most common steps of image-guided interventions (see figure 1). At planning time, a 3D CT scan of the patient is acquired, structures of interest are segmented and the planning of the intervention is performed. At this stage, the pose of the C-arm and the pose of the patient are often chosen. Then, at treatment time, the patient and the C-arm are positioned following the treatment planning. To position the C-arm in the planned pose, a rigid registration must be performed. This step may be performed interactively [13] or automatically [14]. In either case, the C-arm cannot be aligned perfectly to the targeted pose, resulting in a small pose error that must be accounted for by the registration method. Finally, a non-rigid registration method is often needed to compensate for the possible large non-rigid deformations and map the segmented structures of interest onto the fluoroscopic image. Remember that our objective is to recover a 3D displacement field, and not just a mapping from the 3D to the 2D image space.

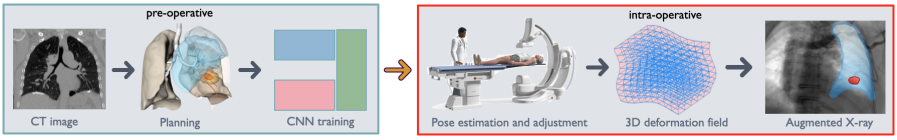

Fig. 1 Overview of our method: first, using a single 3D CT scan of the patient, we plan the intervention (i.e. the anatomical structures of interest are segmented and determine the optimal C-arm pose). Second, we train a neural network using randomized deformations and small randomized rigid transformations. Third, at the time of the intervention, the C-arm is adjusted to the planned pose, an X-ray image is acquired, and the network predicts in real-time the 3D deformation field from which we visually augment the fluoroscopic image.

As discussed in section 1, non-rigid registration methods often assume no change between the C-arm pose at planning time and treatment time. However, because rigid registration methods have a limited accuracy, this assumption is invalid. For example, the mean 2D reprojection error of a state of the art automatic registration method ([14]), is $7.8mm$. Consequently, our framework incorporates pose variation in the data generation process. At training time, the pose variation is composed with the displacement field to produce a target displacement field for the network to learn.

2.1 Network architecture

Our network architecture is inspired by the work of Shen *et al.* [15] which goal is to reconstruct a CT volume from a fluoroscopic image. In this fully convolutional architecture, the direct translation from a dense 2D input to a dense 3D output is a key characteristic to learn fluoroscopy-to-CT mapping.

In order to use this network for 2D-3D registration, the output of the network is now a 3D displacement field representing the total non-rigid + rigid

transform of the preoperative CT to the intra-operative anatomy visible in the fluoroscopic image.

This allows the network to predict both global and local displacements in the preoperative CT from the intra-operative fluoroscopic image. We also added several Adversarial Noise Layers (ANL) [16] to regularize the latent space. The architecture of our network is summarized in figure 2.

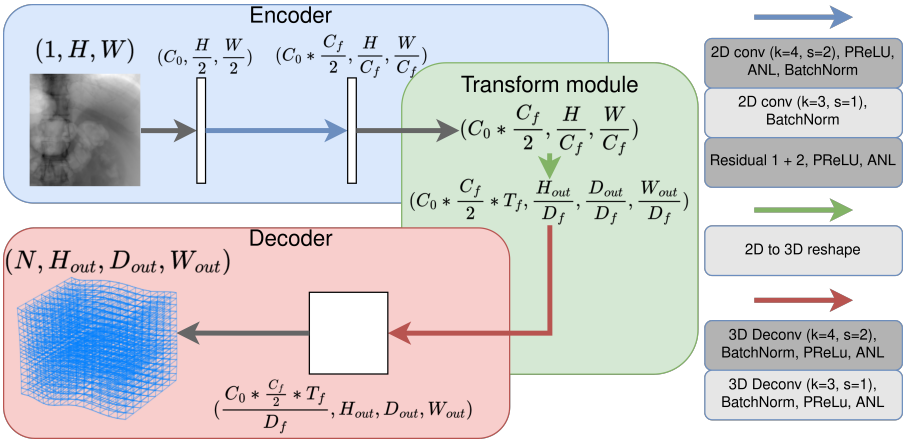

Fig. 2 Our network is based on an encoder-decoder architecture. The abbreviations k and s stand for kernel size and stride, respectively. The network takes as input one fluoroscopic image and outputs a sub-sampled 3D vector field on the CT volume. The central $(N_c - 2)$ layers of the encoder are arranged in Residual Blocks. The data shape is fully described by: $C_f = 2^{N_c/2}$, $D_f = 2^{N_{dc}/2}$, and $T_f = (HWD_f^3)/(H_{out}D_{out}W_{out}C_f^2)$.

The 2D input is an image of size $(1, H, W)$. The output 3D displacement field can also be considered as 3-channel volumetric image, of size $(3, H_{out}, D_{out}, W_{out})$. The first convolutional layer, represented by a gray arrow, with $k = 4$ and $s = 2$, transforms the input into a feature map with C_0 features. For every following Residual Block, H and W are divided by 2 and C_0 is multiplied by 2. The last convolutional layer, in gray, with $k = 3$ and $s = 1$, is not in a residual block and is followed by a BatchNorm, PRelu and ANL. The transform module performs a 2D to 3D reshape of the feature maps extracted by the encoder and passes them to the decoder. In the decoder, H, D and W are upscaled by a factor of 2 while the number of features is divided by 2 for every two of the N_{dc} layers. An additional deconvolutional layer, in gray, with $k = 1$ and $s = 1$, transforms the $\frac{C_0 * C_f}{D_f}$ output features into 2 channels. The 2 channels correspond to the two directions of displacement visible in the projection image. Predicting the displacement component perpendicular to the image plane is very challenging as a deformation in this direction leads to nearly no change in the fluoroscopic image. The network is supervised against the known 3D displacement field using an MSE loss, using AdamW [17] ($\lambda = 0.05$) as the optimizer. An ablation study (see table 1) was performed to determine the best values for the hyperparameters $N_c, N_{dc}, C_0, H, W, H_{out}, W_{out}, D_{out}$.

2.2 Data generation

To train our network, we generate a synthetic dataset composed of DRRs paired with 3D displacement fields, which are used to deform the preoperative CT I_0. From the deformed CT Scans, we generate DRRs using the DeepDRR algorithm [18]. We detail below these different steps and motivate our choices.

Deformation generation:

To obtain smooth and invertible deformations $u(x)$, we generate Displacement Vector Fields (DVFs) following the LDDMM framework [19]. Under this framework, the transformation ϕ that registers an image I_0 to an image I_1 is defined by $\phi(x) = x+u(x)$, of inverse $\phi^{-1}(x) = x-u(x)$, such that $\left\|I_0 \circ \phi^{-1}(x) - I_1\right\|^2$ is minimized. The vector field $u(x)$ is computed by integrating a velocity field $v(t,x)$ over time. A set of differential equations drive the evolution of $v(t,x)$ to satisfy the minimization condition.

In Durrleman *et al.* [20], the authors demonstrate that $v(t,x)$ can be expressed as $v(t,x) = \sum_{k=1}^{N_{cp}} K(x, c_k(t))\alpha_k(t)$. In this formulation, K is a element of a Reproducing Kernel Hilbert Space. From an implementation point of view, this permits to use a Gaussian kernel as K, and the solution of the minimization problem $u(x)$ is obtained in this case through the evolution of α_k and c_k.

Domain randomization:

To implement Domain Randomization [21] for deformation field generation, we randomly sample $\alpha_k(t)$ and $c_k(t)$ to generate a large variety of diffeomorphic displacement fields. This is done by first assigning a uniform probability distribution $U_{\alpha_k}(\mu_{\alpha_k} - \frac{w_\alpha}{2}, \mu_{\alpha_k} + \frac{w_\alpha}{2})$ for each α_k and $U_{c_k}(\mu_{c_k} - \frac{w_c}{2}, \mu_{c_k} + \frac{w_c}{2})$ for each c_k, with $(\mu_{\alpha_k}, \mu_{c_k})$ randomized for each sample. At each time step t, $\alpha_k(t)$ and $c_k(t)$ are sampled from U_{α_k} and U_{c_k} to compute v_t. Because $v_t(x_t)$ is analytical, we can check that $\|v_t(x_t)\|_{W^{1,\text{inf}}} < 1$ where $\|v_t\|_{W^{1,\text{inf},(\mathbb{R}^N,\mathbb{R}^N)}} = sup_{x\in\mathbb{R}^N}(|v_t(x_t)|_{\mathbb{R}^N} + |\nabla v_t(x_t)|_{\mathbb{R}^N\times\mathbb{R}^N})$ which ensures $u_t(x_t) = x_t + v_t(x_t)$ is a diffeomorphism [22] with N the number of control points. We then obtain x_{t+1} by $x_{t+1} = x_t + v_t$ before computing $v_{t+1}(x_{t+1})$. This way $u(x)$ is built iteratively from $t = 0$ to $t = t_{max}$ and we can guarantee that u remains a diffeomorphism. An additional check is performed at each time step to ensure that the volume of each voxel is not reduced below a given threshold v_{thresh}. With this process, we make sure that information is not lost when applying the displacement field to the image, and that the random DVFs remain diffeomorphisms. Through that process, we can apply the principles of domain randomization to our problem, and train the network in a generic way to ensure robustness and unbiasedness towards preferred directions of deformation. The generated DVFs are then applied to the preoperative CT via the *grid_sample* function in PyTorch.

Digitally Reconstructed Radiographs:

Because fluoroscopy involves ionizing radiations, obtaining a sufficient number of real images to train the network is not feasible. Instead, we generate Digitally Reconstructed Radiographs (DRR) using the DeepDRR framework [18]. This framework models the C-arm as a pinhole camera, parameterized by an intrinsic matrix K and an extrinsic matrix $E = \left[\begin{array}{c|c} R & T \\ \hline 0 & 1 \end{array}\right]$ with R a $(3,3)$ rotation matrix and T a $(3,1)$ translation vector. Values for T and R are defined during the surgical planning stage (see figure 1). The intrinsic matrix is defined by detector characteristics, that are fixed for a given C-arm detector panel. The DRR projection p is generated from K, E and the CT volume I_n, as presented in [18].

Pose sampling:

While steps are taken to ensure the pose of the C-arm at treatment time is close to the planning pose, this step is still prone to errors (see 1), which is why the training data must include pose variations. A pose change matrix P is parameterized in the same way as E, by a translation T_P and a rotation R_P. The updated pose is given by $E' = PE$. T_P if first sampled uniformly as a 2D vector parallel to the image plane with an amplitude between 0 and 1. T_P is then scaled by $a \sim \mathcal{N}_{\mathcal{T}}(0, \sqrt{T_{max}/2})$ with T_{max} the amplitude of translations for the dataset. R_P is sampled uniformly from the Haar distribution. R_P is then converted to a rotation vector $r_P \in \mathbb{R}(3\times 1)$. r_P is then scaled in the same way as T_P, by $b \sim \mathcal{N}_{\mathcal{R}}(0, \sqrt{R_{max}/2})$, with R_{max} the amplitude of rotations for the dataset, and converted back to matrix representation. Finally, the tail of the normal distributions for T_{max} and R_{max} were cut for $\|T\| > T_{max}$ and $\|R\| > R_{max}$ to remove outliers. R_P and T_P are sampled normally because the most likely pose of the C-arm at treatment time should be the planning pose, and poses far from the planning pose should be less frequent than poses close to the planning pose.

Data generation summary:

A data sample i is composed of p_i, $u_i(x)$, P_i. First, a displacement field $u_i(x)$ and a deformed CT I'_0 are generated via our domain randomization approach. Our pose sampling process is used to obtain P_i and E'. Finally, knowing I'_0 and E', we generate the projection p_i via DeepDRR.

3 Results & discussion

Experimental setup:

Using a lung 3D CT from the COVID-19-AR dataset [23], we generated 10,000 data samples. These samples were split between a training dataset containing 8,000 samples and a validation dataset containing 2,000 samples.

The data generation process described in 2.2 was employed to generate the training/validation dataset. Assuming a maximum deformation amplitude of

Parameter considered	Validation loss
$(H, W) = (128, 128)$; $(H_{out}, D_{out}, W_{out}) = (32, 16, 32)$	49.22
$(H, W) = (256, 256)$; $(H_{out}, D_{out}, W_{out}) = (64, 32, 64)$	**38.41**
$N_c = 6$	**38.41**
$N_c = 10$	53.69

Table 1 Ablation study reporting the lowest obtained validation loss for different values of $N_c = N_{dc}$, C_0, $H = W$, $H_{out} = W_{out}$, $D_{out} = \frac{H_{out}}{2}$.

30 mm, we then defined the associated Domain Randomization parameters. Thus, each of the 3 components of μ_{α_k} and μ_{c_k} were drawn from uniform distributions $U(0, 10)$ and $U(12, 72)$. The values of w_α and w_c were set to 0.001 mm and 0.6 mm respectively. Finally, the threshold v_{thresh} was set to $0.75mm^3$ and the displacement field was generated in $t_{max} = 100$ iterations. For the DRR generation, the distance from the detector to the volume center was set to $\|T\| = 1500$ mm, and R was defined such that the image plane normal was co-linear to the AP axis of the volume. Finally, the parameters used for pose sampling were $T_{max} = 17$ mm and $R_{max} = \frac{\pi}{4}$. To predict the total rigid + non-rigid transformation, the rigid transform was applied to a regular grid of points deformed by the non-rigid transform to obtain the final displacement field u_{tot}. This displacement field was used as the target displacement field for the network.

The network was trained for 75 epochs with an initial learning rate of 10^{-4} decreased to 10^{-5} after 30 epochs. Training took about 26 hours on a computer equipped with an AMD Ryzen 3950X CPU and an Nvidia Titan RTX GPU. The memory footprint of the network remains small, at 220 MB.

An ablation study was performed to determine the optimal network hyperparameters (see table 1). Each hyperparameter was varied independently while the other parameters were fixed to their best values. The optimal hyperparameters are $N_c = 6, C_0 = 64, H_{out} = 64$. Some values combinations could not be tested due to memory constraints, eg. $H = 512, C_0 = 64, H_{out} = 64, N_c = 6$.

Results:

The accuracy of the network was evaluated on 6 anatomical landmarks covering both lungs. We measured the (3D) TRE between the predicted and ground truth positions but also the projection distance (PD) as suggested in [24].

On the validation dataset containing pose variations and domain randomized synthetic deformations, we obtained a mean TRE (mTRE) of 2.26±2.00 mm and a PD of 7.72±5.00 mm. In this validation dataset, landmark displacements range from 0.83 mm to 27.77 mm in 3D and 2.86 to 95 mm in 2D. This first result shows that our method is on par with the state of art even for large amplitudes of displacement while also accounting for pose errors.

To further validate our method, we tested it on a series of deformations from a 10-phase respiratory-correlated lung 4DCT provided by the 4D-Lung

Phase #	mean TRE (mm)
0	0.8 ± 0.4
1	1.85 ± 1.28
2	3.02 ± 1.6
3	2.97 ± 2.57
4	2.96 ± 1.21
5	2.76 ± 1.67
6	3.16 ± 1.09
7	1.97 ± 0.84
8	1.84 ± 0.93
9	1.71 ± 0.53

Table 2 Prediction results on 10 phases of a breathing cycle with only non-rigid displacements.

Phase #	mean TRE (mm)	mean PD (mm)
0	4.48 ± 4.41	1.47 ± 0.88
1	3.44 ± 2.74	1.4 ± 0.85
2	4.85 ± 4.49	1.95 ± 1.27
3	5.12 ± 5.07	2.0 ± 1.42
4	5.03 ± 3.36	2.21 ± 1.34
5	4.29 ± 2.61	2.31±1.44
6	3.97 ± 3.07	1.87±1.12
7	4.16 ± 2.85	2.04±1.25
8	5.48 ± 3.88	2.1±1.64
9	3.66 ± 3.17	1.55±1.14

Table 3 Prediction results on 10 phases of a breathing cycle when both rigid and non-rigid displacements are present.

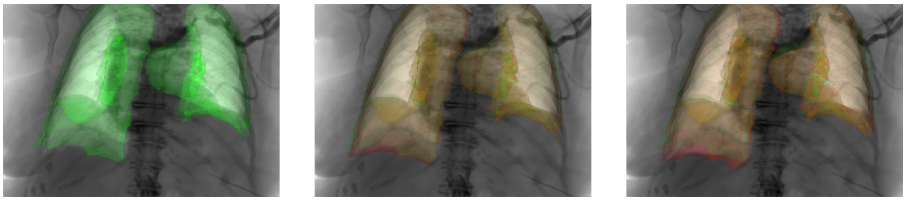

Fig. 3 Illustration of a lung segmentation deformed to match breathing motion, in red, superposed with the pre-operative segmentation, in green. The segmentations are overlayed on top of a DRR, simulating an augmented fluoroscopy.

dataset [25]. For the first testing dataset, the pose is not varied and the network is trained to recover only a non-rigid displacement. For the second testing dataset, 20 poses were sampled for each of the 10 phases of the 4DCT to generate a testing dataset containing 200 samples. The results for the first and second datasets are summarized in table 2 and table 3, respectively. We can see from the table 2 that the domain randomization approach for displacement fields generation is justified, as the network performs very well (mTRE < 3.16 mm) for realistic deformations even though it was trained on synthetic data. However, assuming that the pose is exactly known before the intervention is not realistic, hence the need to account for pose uncertainty. The table 3 presents the results for a more challenging case, where the inputs are generated with a varying pose ($T_{max} = 17mm$ and $R_{max} = \frac{\pi}{4}$). The network has thus been trained to recover such pose changes, and the accuracy remains high (mTRE < 5.48 mm, mPD < 2.31 mm) even though the 2D-3D registration task is now to estimate a rigid + non-rigid transform.

4 Conclusion

The objective of our work was to propose an accurate and real-time method able to recover a 3D displacement field from a single 2D fluoroscopic image. We show that this ill-posed problem can be solved via modern deep learning techniques when associated with a comprehensive data generation pipeline.

Our method only requires routinely acquired images (a single preoperative 3D CT and a single projection X-ray image at test time), and is robust to variations in pose, making it applicable to actual clinical scenarios. Our results show that the proposed method can estimate a 3D displacement field, even for structures deep into the tissues, with an average accuracy of 4.45 mm.

This level of accuracy is obtained at an update rate of about 20 Hz,

sufficient for interactive visualization of an augmented fluoroscopic image as illustrated in figure 3, or for real-time 3D navigation.

Our next steps will focus on the estimation of displacements that are perpendicular to the image plane, possibly through the use of a biomechanical model.

Acknowledgments. The authors would like to thank Juan Verde, MD and Simon Rouze, MD for their valuable inputs during the development of this method. This work was funded by the French national research agency ANR (VATSOP project).

References

[1] Sharma, M., Nano, T.F., Akkati, M., Milano, M.T., Morin, O., Feng, M.: A systematic review and meta-analysis of liver tumor position variability during sbrt using various motion management and igrt strategies. Radiotherapy and Oncology **166**, 195–202 (2022)

[2] Unberath, M., Gao, C., Hu, Y., Judish, M., Taylor, R., Armand, M., Grupp, R., Kwok, K., Manfredi, L., Li, C.: The Impact of Machine Learning on 2D/3D Registration for Image-Guided Interventions: A Systematic Review and Perspective. Frontiers in Robotics and AI **8** (2021)

[3] Sotiras, A., Davatzikos, C., Paragios, N.: Deformable Medical Image Registration: A Survey. IEEE Transactions on Medical Imaging (2010)

[4] Kilburn, J.M., Soike, M.H., Lucas, J.T., Ayala-Peacock, D., Blackstock, W., Isom, S., Kearns, W.T., Hinson, W.H., Miller, A.A., Petty, W.J., Munley, M.T., Urbanic, J.J.: Image guided radiation therapy may result in improved local control in locally advanced lung cancer patients. Practical Radiation Oncology **6**(3), 73–80 (2016)

[5] Adler, J.R., Chang, S.D., Murphy, M.J., Doty, J., Geis, P., Hancock, S.L.: The Cyberknife: A frameless robotic system for radiosurgery. In: Stereotactic and Functional Neurosurgery, vol. 69, pp. 124–128 (1997)

[6] Seppenwoolde, Y., Wunderink, W., Mc, E., Romero, A.M.: Treatment precision of image-guided liver SBRT using implanted fiducial markers depends on marker-tumour distance. Physics in Medicine and Biology (2011)

[7] Suzuki, K., Nagai, K., Yoshida, J., Ohmatsu, H., Takahashi, K., Nishimura, M., Nishiwaki, Y.: Video-assisted thoracoscopic surgery for small indeterminate pulmonary nodules. Chest **115**(2), 563–568 (1999)

[8] Keating, J., Singhal, S.: Novel methods of intraoperative localization and margin assessment of pulmonary nodules. Seminars in Thoracic and Cardiovascular Surgery **28**(1), 127–136 (2016)

[9] Shieh, C.C., Caillet, V., Dunbar, M., Keall, P.J., Booth, J.T., Hardcastle, N., Haddad, C., Eade, T., Feain, I.: A Bayesian approach for three-dimensional markerless tumor tracking using kV imaging during lung radiotherapy. Physics in Medicine and Biology **62**(8), 3065–3080 (2017)

[10] Zhang, Y., Huang, X., Wang, J., Sebastian, N., Robb, R., Webb, A., Shilo, K., Denicola, G.M., Williams, T.M.: Automatic Cone Beam Projection-based Liver Tumor Localization by Deep Learning and Biomechanical Modeling. Int. Journal of Radiation Oncology, Biology, Physics **108**(3), 171 (2020)

[11] Hirai, R., Sakata, Y., Tanizawa, A., Mori, S.: Real-time tumor tracking using fluoroscopic imaging with deep neural network analysis. Physica Medica **59**, 22–29 (2019)

[12] Foote, M.D., Zimmerman, B.E., Sawant, A., Joshi, S.C.: Real-Time 2D-3D Deformable Registration with Deep Learning and Application to Lung Radiotherapy Targeting. Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics) **11492 LNCS**, 265–276 (2019)

[13] Rouze, S., de Latour, B., Flecher, E., Guihaire, J., Castro, M., Corre, R., Haigron, P., Verhoye, J.-P.: Small pulmonary nodule localization with cone beam computed tomography during video-assisted thoracic surgery: a feasibility study. Interactive CardioVascular and Thoracic Surgery **22**(6), 705–711 (2016)

[14] Lee, B.C., Sinha, A., Varble, N., Pritchard, W.F., Karanian, J.W., Wood, B.J., Bydlon, T.: Breathing-compensated neural networks for real time c-arm pose estimation in lung ct-fluoroscopy registration. In: IEEE International Symposium on Biomedical Imaging (ISBI), pp. 1–5 (2022)

[15] Shen, L., Zhao, W., Xing, L.: Patient-specific reconstruction of volumetric computed tomography images from a single projection view via deep

learning. Nature Biomedical Engineering **3**(11), 880–888 (2019)

[16] You, Z., Ye, J., Li, K., Xu, Z., Wang, P.: Adversarial Noise Layer: Regularize Neural Network by Adding Noise; Adversarial Noise Layer: Regularize Neural Network by Adding Noise. 2019 IEEE International Conference on Image Processing (ICIP) (2019)

[17] Loshchilov, I., Hutter, F.: Decoupled weight decay regularization. In: International Conference on Learning Representations (2018)

[18] Unberath, M., Zaech, J.N., Lee, S.C., Bier, B., Fotouhi, J., Armand, M., Navab, N.: DeepDRR – A Catalyst for Machine Learning in Fluoroscopy-Guided Procedures. Lecture Notes in Computer Science **11073 LNCS**, 98–106 (2018). Accessed 2022-02-27

[19] Trouve, A., Faisal Beg, M., Miller, M.I., Younes, L.: Computing Large Deformation Metric Mappings via Geodesic Flows of Diffeomorphisms. International Journal of Computer Vision **61**(2), 139–157 (2005)

[20] Durrleman, S., Prastawa, M., Charon, N., Korenberg, J.R., Joshi, S., Gerig, G., Trouvé, A.: Morphometry of anatomical shape complexes with dense deformations and sparse parameters. NeuroImage **101**, 35–49 (2014)

[21] Tobin, J., Fong, R., Ray, A., Schneider, J., Zaremba, W., Abbeel, P.: Domain Randomization for Transferring Deep Neural Networks from Simulation to the Real World. In: IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), pp. 22–30 (2017)

[22] Banyaga, A.: The Structure of Classical Diffeomorphism Groups. Mathematics and its Applications, vol. 400. Kluwer Academic, ??? (1997)

[23] Shivang Desai, e.a.: Chest imaging representing a COVID-19 positive rural u.s. population. Scientific Data **7**(1) (2020)

[24] van de Kraats, E.B., Penney, G.P., Tomazevic, D., van Walsum, T., Niessen, W.J.: Standardized evaluation methodology for 2-d-3-d registration. IEEE Transactions on Medical Imaging **24**(9), 1177–1189 (2005)

[25] Hugo, G.D., Weiss, E., Sleeman, W.C., Balik, S., Keall, P.J., Lu, J., Williamson, J.F.: A longitudinal four-dimensional computed tomography and cone beam computed tomography dataset for image-guided radiation therapy research in lung cancer. Medical physics **44**(2), 762 (2017)