Evaluation of Multi-metric Registration for Online Adaptive Proton Therapy of Prostate Cancer

Mohamed S. Elmahdy[1(✉)], Thyrza Jagt[3], Sahar Yousefi[1], Hessam Sokooti[1], Roel Zinkstok[1], Mischa Hoogeman[3], and Marius Staring[1,2]

[1] Leiden University Medical Center, Leiden, The Netherlands
m.s.e.elmahdy@lumc.nl
[2] Delft University of Technology, Delft, The Netherlands
[3] Erasmus MC Cancer Institute, Rotterdam, The Netherlands

Abstract. Delineation of the target volume and Organs-At-Risk (OARs) is a crucial step for proton therapy dose planning of prostate cancer. Adaptive proton therapy mandates automatic delineation, as manual delineation is too time consuming while it should be fast and robust. In this study, we propose an accurate and robust automatic propagation of the delineations from the planning CT to the daily CT by means of Deformable Image Registration (DIR). The proposed algorithm is a multi-metric DIR method that jointly optimizes the registration of the bladder contours and CT images. A 3D Dilated Convolutional Neural Network (DCNN) was trained for automatic bladder segmentation of the daily CT. The network was trained and tested on prostate data of 18 patients, each having 7 to 10 daily CT scans. The network achieved a Dice Similarity Coefficient (DSC) of $92.7\% \pm 1.6\%$ for automatic bladder segmentation. For the automatic contour propagation of the prostate, lymph nodes, and seminal vesicles, the system achieved a DSC of 0.87 ± 0.03, 0.89 ± 0.02, and 0.67 ± 0.11 and Mean Surface Distance of 1.4 ± 0.30 mm, 1.4 ± 0.29 mm, and 1.5 ± 0.37 mm, respectively. The proposed algorithm is therefore very promising for clinical implementation in the context of online adaptive proton therapy of prostate cancer.

Keywords: Deformable image registration
Convolutional neural networks (CNN) · Prostate cancer
Proton therapy

1 Introduction

Prostate cancer is one of the leading causes of mortality and the most common cancer among men. The American Cancer Society estimates around 161,360 new cases and 26,730 deaths from prostate cancer in the United States for 2017 [1]. Intensity-Modulated Proton Therapy (IMPT) has shown the capability of delivering a highly localized dose distributions to the target volume. IMPT is

however more sensitive to daily variations that may result in a suboptimal dose distribution [2,3]. These variations could be due to anatomical changes in the target volume and Organs-At-Risk (OARs) or a misalignment in the patient positioning. In order to account for these variations, a margin is added to the Clinical Target Volume (CTV) that leads to the Planning Target Volume (PTV). These margins result in extra dose to the OARs, leading to an increase in the treatment-related toxicities that may prevent dose escalation. Repeat imaging and re-planning can handle this problem [4]. These repeat (inter-fraction) CT scans have to be delineated first before generating a new treatment plan. Therefore traditionally the inter-fraction re-contouring is not performed because it is very time consuming and consequently new anatomical changes could be introduced in the meantime. Therefore, it is vital for the automatic contouring to be fast and robust, because otherwise there will be a need for fallback strategies like manual correction that also take time.

The Atlas Based Auto Segmentation (ABAS) tool, Mirada, RayStation, and MIM softwares are a well known commercial softwares for automatic re-contouring. However, these softwares are considered a black box for the users, and therefore limits the potential of parameter customization and tuning. Open source DIR packages provide a high level of flexibility with a concrete scientific evidence and reproducibility [5,6]. Qiao *et al.* [7] reported an MSD of 1.36 ± 0.30 mm, 1.75 ± 0.84 mm, 1.49 ± 0.44 mm for the prostate, seminal vesicles, and lymph nodes, respectively for 18 patients using the open source `elastix` software. A clinical success rate of 69% was achieved, which means that 31% of the delineations have to be corrected, leading to increased costs and a suboptimal patient workflow. In 2011, Thor *et al.* deployed DIR to propagate the contours of the prostate and OARs from CT to cone-beam CT [8]. The system achieved a mean DSC of 0.80 for prostate, 0.77 for rectum, and 0.73 for the bladder with a relatively high variance. Moreover, the system was not qualitatively evaluated in terms of the dosimetric coverage. Recently, Woerner *et al.* [9] investigated the error between different radiologists and both DIR and rigid registration in different body regions. They only reported the results for the prostate, which were 0.90, 0.99 mm, and 8.12 mm for the DSC, MSD, and Hausdorff Distance (HD), respectively.

In order to improve the success rate of the automatic propagation of contours using DIR, we propose a multi-metric based registration. Hereby, instead of depending on the intensity of the images alone, we introduce a second objective that specifically optimizes the bladder overlap, based on a bladder estimate provided by a neural network.

2   Materials and Methods

2.1   Dataset

This study includes eighteen anonymized patients who were treated for prostate cancer in 2007 using intensity-modulated radiation therapy at Haukeland University Hospital. Each patient has a planning CT and 7 to 10 repeat CT scans.
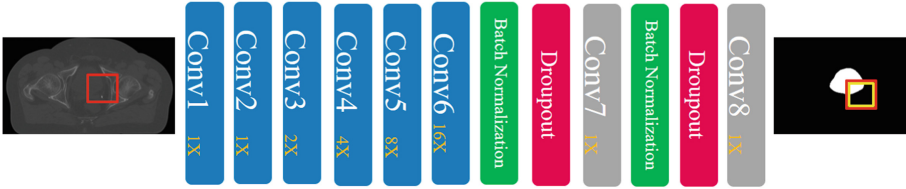
Fig. 1. The architecture for the 3D-DCNN network where $\{1X, \ldots, 16X\}$ denotes the dilation rate. The blue convolution blocks represent $3 \times 3 \times 3$ kernels while the grey convolution blocks represent fully connected convolution layers implemented by $1 \times 1 \times 1$ kernels. The green and red blocks denote batch normalization layers and dropout layers, respectively. The red square represents the patch, while the yellow square represents the receptive field. (Color figure online)

The field of view of the scans included the prostate, lymph nodes, seminal vesicles, in addition to the bladder and rectum as the main OARs. Each scan has 90 to 180 slices with a slice thickness of around 2 mm. All the slices were of size 512×512 with in-plane resolution of around 0.9 mm. The prostate, lymph nodes, seminal vesicles, rectum, and bladder were delineated in each CT scan by an expert and reviewed by another one.

2.2   Dilated Convolution Neural Network Architecture (DCNN)

Motion and filling of the bladder as well as the rectum have an important influence on the anatomical changes in the abdomen. Therefore, we hypothesize that intensity-based DIR may improve in terms of accuracy and robustness if the motion of either of these structures is taken into account explicitly. Since the bladder is a well-defined structure that is relatively easy to delineate, we opt to segment it fully automatically. In this study, we propose a 3D Dilated Convolutional Neural Network (3D-DCNN) in order to automatically segment the bladder. Dilated convolution is a generalized version of the traditional convolution process where more spacing is added to the convolution kernel so that a larger spatial neighborhood is considered when calculating the feature maps. This spacing is called the dilation rate; for traditional convolution the dilation rate is 1. Using a dilation rate larger than 1 has several advantages. First, stacking convolution layers with increasing dilation rate will accordingly enlarge the Receptive Field (RF) of the neural network without adding additional trainable parameters. Second, there is no need for adding down-sampling layers to have a large RF and therefore the network can handle high resolution volumes using a smaller number of trainable parameters. Figure 1 shows the architecture of the dilated network. This network is a modified version of the architecture deployed in [10]. The first six convolutional layers have a kernel size of $3 \times 3 \times 3$, 32 feature maps, and a logarithmic increasing dilation rate. Dropout layers with a dropout rate of 0.6 as well as batch normalization layers are introduced before the last two fully convolutional layers. Moreover, the 2D convolution layers in
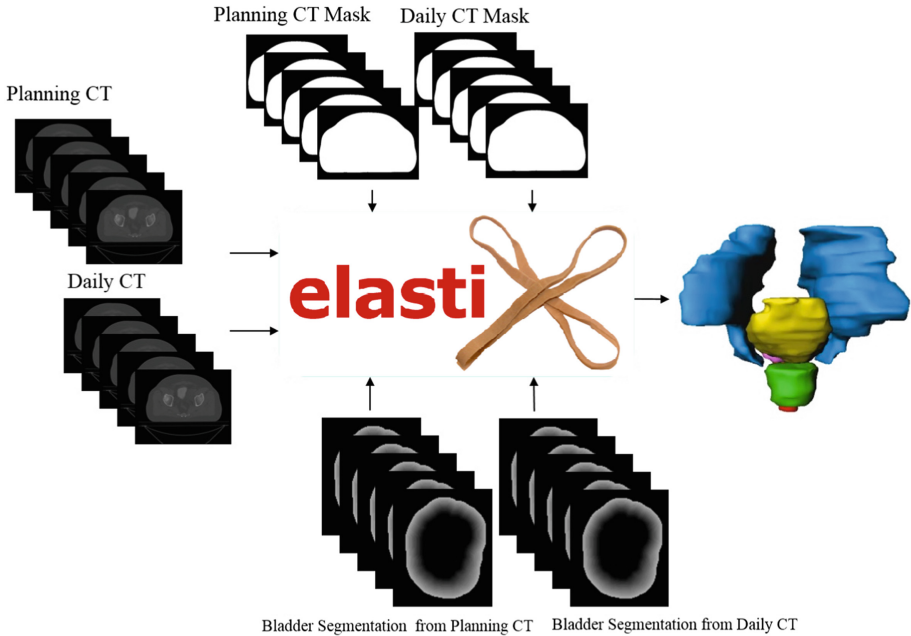
Fig. 2. The proposed multi-metric registration process using `elastix` software.

the original architecture were replaced with 3D layers in order to consider the homogeneity of tissues in 3D. Hence, it can help to get more accurate and robust results. The network has a receptive field of $65 \times 65 \times 65$ and has 144,551 trainable parameters.

In order to train the network, the 18 patients are divided into three sets: 12 patients for training, 3 patients for validation, and 3 patients for testing. This results in a total of 120, 28, and 30 CT scans for training, validation, and testing, respectively. 1,000,000 patches of size $71 \times 71 \times 71$ are randomly extracted from the training volumes, making sure they are equally distributed between foreground and background. Cross Entropy is deployed as a cost function and the network is trained using the Adam optimizer with a fixed learning rate of 10^{-4}. All the experiments were carried out using an NVIDIA GTX1080 Ti with 11 GB of GPU memory.

2.3 Image Registration

The open source package `elastix` was used for deformable image registration [6]. This package is available from http://elastix.isi.uu.nl. All the experiments were performed on a desktop PC operated on Windows 10 with 16 GB of memory and an Intel Xeon E51620 CPU with 4 cores running at 3.6 GHz.

In this study, the planning CT of each patient together with the manual delineation of the bladder are considered the moving images, while the repeat CT of the same patient accompanied with the bladder segmentation resulting from the proposed 3D-DCNN are the fixed images. The generated Deformation Vector Field (DVF) is then used to propagate the remaining contours (prostate, seminal vesicles, lymph nodes, and rectum) from the planning CT to the repeat CT. In order to have a good registration initialization, the registrations were initialized based on the center-of-gravity of the bony anatomy defined by a Hounsfield number larger than 200. To remove the effect of the CT table, a mask of the body torso was generated using Pulmo software [11]. The registration process is a two step procedure. First, the CT images are aligned using a single resolution affine transformation so that we can eliminate the large organ movements. Second, a deformable registration is applied to tackle the local deformations of the organs. A fast recursive implementation of the B-spline transformation was employed for DIR [12]. Adaptive stochastic gradient descent was used for optimization [13]. Figure 2 illustrates the proposed registration pipeline. For the DIR stage we used a three level Gaussian pyramid, and two cost functions. Mutual information was used for the CT images. To take into account the bladder contours the distance transform of the bladder segmentations is used instead of the binary segmentations themselves, to ensure a smooth and stable optimization process. This results in the following optimization problem:

$$\hat{\mu} = \arg\min_{\mu}\{C_1\left(I_F, I_M; T_\mu\right) + \alpha C_2\left(DT\left(S_F\right), DT\left(S_M\right); T_\mu\right)\}, \qquad (1)$$

where C_1 is the mutual information cost function, C_2 is the Mean Square Difference (MSD) cost function, α is a weight balancing these two cost functions, I_F is the daily scan, I_M is the planning scan, $DT\left(S_F\right)$ is the distance transform of the DCNN bladder segmentation, and $DT\left(S_M\right)$ is the distance transform of the manual annotation of the planning scan.

2.4   Registration Performance Evaluation

The registration quality is measured by the overlap and residual distance between the manually and the automatically propagated contours of the daily CT for the prostate, lymph nodes, seminal vesicles, rectum, and bladder. The most common metrics for quality are the Dice Similarity Coefficient (DSC), the Mean Surface Distance (MSD), and the Hausdorff Distance (HD), all computed in 3D.

$$DSC = \sum \frac{2|F \cap M|}{|F| + |M|}, \qquad (2)$$

where F and M are the propagated contour and the ground truth contour, respectively.

$$MSD = \frac{1}{2}\left(\frac{1}{n}\sum_{i=1}^{n} d\left(a_i, M\right) + \frac{1}{m}\sum_{i=1}^{m} d\left(b_i, F\right)\right), \qquad (3)$$

$$HD = \max \left\{ \max_i \left\{ d\left(a_i, M\right) \right\}, \max_j \left\{ d\left(b_i, F\right) \right\} \right\}, \tag{4}$$

where $\{a_1, a_2, ..., a_n\}$ and $\{b_1, b_2, ..., b_m\}$ are the surface mesh points of the fixed and moving contours, respectively and $d\left(a_i, M\right) = \min_j ||b_j - a_i||$.

3   Experimental Results

3.1   DCNN Segmentation Performance

The DCNN network achieved an average segmentation DSC of $92.7\% \pm 1.6\%$ on the test patients. It took an average of 15 s to segment a single volume using a single GPU depending on the number of slices per volume.

3.2   Registration Performance

The weight α of the cost function for the bladder segmentation (C_2) was set to 0.05 for the first resolution and zero for the second and third resolutions. These weights were chosen after a set of initial experiments. For investigating the effect of the number of iterations on the registration performance, we varied this parameter between 100 and 500 iterations. Table 1 illustrates the DSC evaluations of the single-metric and multi-metric registrations for the set of iterations. The overlap performance of the prostate, lymph nodes, and rectum were very similar for single and multi metric registrations. For the seminal vesicles and bladder the overlap was higher for multi-metric at 100 and 500 iterations.

The evaluations in terms of MSD are shown in Table 2. For the prostate, seminal vesicles, rectum, and bladder, there was a significant improvement from the affine transformation to DIR-100, and a slight improvement for 500 iterations in both single and multi-metric registrations. This was not the case for lymph nodes. However, the MSD errors for almost all the organs were within a voxel size. The 95% HD showed a similar pattern as MSD as presented in Table 3.

Figure 3 shows the comparison of the registration performance between single-metric (intensity image only) and multi-metric registrations (intensity and bladder segmentation) for affine, 100, and 500 iterations. The comparison illustrates the performance in terms of DSC, MSD, and 95%HD for the target volumes and OARs. The figure shows much less outliers for the multi-metric registrations, especially for the seminal vesicles, which is a challenging structure due to its small volume. Here, results above the top whisker (defined by 1.5 times the inter-quartile range) are termed an outlier. In order to explore the upper limit of the proposed method, it was tested with the manual annotation of the bladder instead of the segmentation of the DCNN. The boxplot shows a very similar pattern between the multi-metric registration using the bladder contours from the DCNN network and the manually annotated bladder contours.

Table 1. DSC value of the target volumes and OARs for different registration settings. † and ‡ represent a significant difference (at $p = 0.05$) between single-metric and multi-metric for 100 and 500 iterations, respectively.

Evaluation	# it.	Prostate	SV†	LN	Rectum	Bladder†‡
		$\mu \pm \sigma$	$\mu \pm \sigma$	$\mu \pm \sigma$	$\mu \pm \sigma$	$\mu \pm \sigma$
Affine		0.83 ± 0.08	0.34 ± 0.27	0.92 ± 0.03	0.69 ± 0.09	0.78 ± 0.07
Single-metric	100	0.87 ± 0.03	0.59 ± 0.22	0.90 ± 0.02	0.77 ± 0.08	0.90 ± 0.05
	500	0.87 ± 0.04	0.63 ± 0.20	0.89 ± 0.02	0.78 ± 0.07	0.91 ± 0.06
Multi-metric	100	0.87 ± 0.03	0.67 ± 0.11	0.89 ± 0.02	0.78 ± 0.06	0.93 ± 0.03
	500	0.87 ± 0.02	0.66 ± 0.11	0.89 ± 0.02	0.79 ± 0.06	0.93 ± 0.03

Table 2. MSD (mm) of the target volumes and OARs for different registration settings. † and ‡ represent a significant difference (at $p = 0.05$) between single-metric and multi-metric for 100 and 500 iterations, respectively.

Evaluation	# it.	Prostate	SV†	LN	Rectum	Bladder†‡
		$\mu \pm \sigma$	$\mu \pm \sigma$	$\mu \pm \sigma$	$\mu \pm \sigma$	$\mu \pm \sigma$
Affine		1.8 ± 0.78	3.7 ± 2.00	1.1 ± 0.37	4.1 ± 1.50	4.3 ± 1.70
Single-metric	100	1.4 ± 0.33	2.1 ± 1.40	1.3 ± 0.28	3.1 ± 1.30	2.0 ± 1.00
	500	1.3 ± 0.35	1.9 ± 1.40	1.4 ± 0.27	3.0 ± 1.20	1.7 ± 0.87
Multi-metric	100	1.4 ± 0.30	1.5 ± 0.37	1.4 ± 0.29	2.9 ± 0.95	1.4 ± 0.38
	500	1.3 ± 0.28	1.6 ± 0.42	1.4 ± 0.29	2.8 ± 0.88	1.3 ± 0.32

Table 3. %95HD (mm) of the target volumes and OARs for different registration settings. † and ‡ represent a significant difference (at $p = 0.05$) between single-metric and multi-metric for 100 and 500 iterations, respectively.

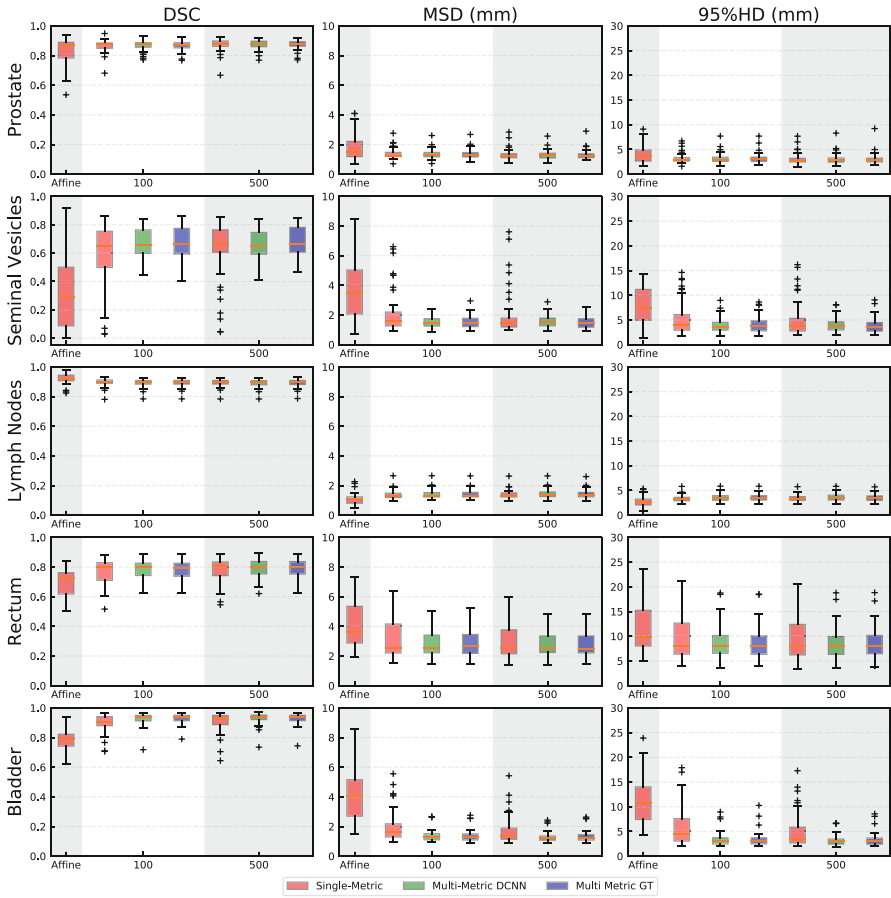| Evaluation	# it.	Prostate	SV†‡	LN	Rectum	Bladder†‡
		$\mu \pm \sigma$	$\mu \pm \sigma$	$\mu \pm \sigma$	$\mu \pm \sigma$	$\mu \pm \sigma$
Affine		4.0 ± 1.70	7.8 ± 3.7	2.7 ± 1.00	11.0 ± 4.7	11.0 ± 4.7
Single-metric	100	3.2 ± 0.96	5.2 ± 3.3	3.3 ± 0.63	9.5 ± 4.3	5.9 ± 3.9
	500	3.1 ± 1.00	4.9 ± 3.4	3.4 ± 0.63	9.3 ± 4.2	5.0 ± 3.4
Multi-metric	100	3.2 ± 0.97	4.0 ± 1.5	3.6 ± 0.71	8.7 ± 3.4	3.4 ± 1.4
	500	3.0 ± 1.00	4.1 ± 1.5	3.6 ± 0.71	8.5 ± 3.2	3.2 ± 1.1

Fig. 3. Boxplot comparison between single-metric and multi-metric image registration versus the number of iterations. The columns show the DSC, MSD, and 95%HD from left to right. Prostate, seminal vesicles, lymph nodes, rectum, and bladder are shown from top to bottom rows, respectively. Here multi-metric DCNN is the result of using the bladder segmentation of the network, while multi-metric GT is the result of using the ground truth bladder delineation.

4   Discussion and Conclusion

In this study, we investigated the hypothesis of enhancing the performance and robustness of the automatic contouring of the target volumes and OARs for prostate cancer using multi-metric Deformable Image Registration (DIR). The purpose of adaptive IMPT is to be able to use a small margin between PTV and CTV, which is only a viable option if the daily re-planning can be performed in an accurate and robust manner. This daily re-planning requires robust automatic re-contouring in order to avoid local treatment-related toxicities and subsequent

adverse side effects. The proposed automatic contouring algorithm was evaluated geometrically. In order to improve the robustness of the registration process, we introduced a multi-metric optimization. This optimization depends not only on the intensity image but also on the segmentation of another organ. In this study, we chose the bladder due to its well defined borders which eases the segmentation process. The quality of the bladder segmentation has a significant effect on steering the registration process, therefore it has to be accurate and robust, so we chose 3D-DCNN. The network achieved a higher DSC than the reported DSC of 81.9% in [14], where a CNN was combined with level sets to segment the bladder in CT urography. It also outperformed the dice overlap of 72% reported in [15], where they attempted to segment all the abdominal organs using a 2D Fully Convolutional Neural Network.

Initializing the registration process using the bony anatomy improved the stability of the registration which is consistent with the findings in [13]. Moreover, introducing a small weighting (α) of 0.05 at the first resolution managed to steer the registration to a better local minima without causing any overfitting to the bladder, therefore there was no need for further weighting in the second and third resolutions.

In this study, we focused on the registration robustness represented by the number of outliers and the variance in the system performance. Overall, the multi-metric registration showed a significant decrease in the number of outliers compared to the single-metric registration. Reducing the number of outliers for the seminal vesicles, which is an important target volume, means a more precise targeting with potential benefits in terms of local control (lower probability of recurrences). Moreover, much less outliers for rectum and especially bladder, which are OARs, means less dose to the OARs with potential benefits in terms of treatment-induced complications after the therapy, so higher probability of better quality-of-life after treatment, see Fig. 3. It also showed a significant improvement in terms of the DSC, MSD, and 95% HD for the seminal vesicles and bladder. For multi-metric registration, the overall performance gets slightly better for 500 iterations and remarkably increased from affine transformation. The figure shows a similar pattern between the multi-metric registration using the manually annotated contours of the bladder and the contours from the DCNN network. This pattern emphasizes that, the proposed method achieved the upper limit of the system. For most of the organs, the registration performance in terms of the MSD was less than 2 mm, which is less than the slice thickness.

In this study, we illustrated the effectiveness of deploying multi-metric registration using the `elastix` software in order to automatically re-contour daily CT scans of the prostate. This re-contouring showed a promise for generating daily treatment plans. Moreover, it showed a substantial improvement in the system robustness, which means that more treatment plans can be directly used without manual correction, which is a crucial factor for enabling online daily adaptation and thus the use of relatively small treatment margins. Therefore, the proposed method could facilitate online adaptive proton therapy of prostate cancer.

References

1. American Cancer Society. https://www.cancer.org/cancer/prostate-cancer/about/key-statistics.html
2. Zhang, M., Westerly, D., Mackie, T.: Introducing an on-line adaptive procedure for prostate image guided intensity modulate proton therapy. Phys. Med. Biol. **56**(15), 4947–4965 (2011)
3. Lomax, A.: Intensity modulated proton therapy and its sensitivity to treatment uncertainties 1: the potential effects of calculational uncertainties. Phys. Med. Biol. **53**(4), 1027–1042 (2008)
4. Hansen, E., Bucci, M., Quivey, J., Weinberg, V., Xia, P.: Repeat CT imaging and replanning during the course of IMRT for head-and-neck cancer. Int. J. Radiat. Oncol. Biol. Phys. **64**(2), 355–362 (2006)
5. Yang, D., Brame, S., El Naqa, I., Aditya, A., Wu, Y., Murty Goddu, S., Mutic, S., Deasy, J., Low, D.: Technical note: DIRART- a software suite for deformable image registration and adaptive radiotherapy research. Med. Phys. **38**(1), 67–77 (2010)
6. Klein, S., Staring, M., Murphy, K., Viergever, M., Pluim, J.: elastix: a toolbox for intensity-based medical image registration. IEEE Trans. Med. Imaging **29**(1), 196–205 (2010)
7. Qiao, Y.: Fast optimization methods for image registration in adaptive radiation therapy. Ph.D. thesis, Leiden University Medical Center (2017)
8. Thor, M., Petersen, J., Bentzen, L., HØyer, M., Muren, L.: Deformable image registration for contour propagation from CT to cone-beam CT scans in radiotherapy of prostate cancer. Acta Oncologica **50**(6), 918–925 (2011)
9. Woerner, A., Choi, M., Harkenrider, M., Roeske, J., Surucu, M.: Evaluation of deformable image registration-based contour propagation from planning CT to cone-beam CT. Technol. Cancer Res. Treat. **16**(6), 801–810 (2017)
10. Wolterink, J.M., Leiner, T., Viergever, M.A., Išgum, I.: Dilated convolutional neural networks for cardiovascular MR segmentation in congenital heart disease. In: Zuluaga, M.A., Bhatia, K., Kainz, B., Moghari, M.H., Pace, D.F. (eds.) RAMBO/HVSMR -2016. LNCS, vol. 10129, pp. 95–102. Springer, Cham (2017). https://doi.org/10.1007/978-3-319-52280-7_9
11. Staring, M., Bakker, M., Stolk, J., Shamonin, D., Reiber, J., Stoel, B.: Towards local progression estimation of pulmonary emphysema using CT. Med. Phys. **41**(2), 021905 (2014)
12. Huizinga, W., Klein, S., Poot, D.H.J.: Fast multidimensional B-spline interpolation using template metaprogramming. In: Ourselin, S., Modat, M. (eds.) WBIR 2014. LNCS, vol. 8545, pp. 11–20. Springer, Cham (2014). https://doi.org/10.1007/978-3-319-08554-8_2
13. Qiao, Y., van Lew, B., Lelieveldt, B.P.F., Staring, M.: Fast automatic step size estimation for gradient descent optimization of image registration. IEEE Trans. Med. Imaging **35**(2), 391–403 (2016)

14. Cha, K.H., Hadjiiski, L., Samala, R.K., Chan, H.-P., Caoili, E.M., Cohan, R.H.: Urinary bladder segmentation in CT urography using deep-learning convolutional neural network and level sets. Med. Phys. **43**(4), 1882–1896 (2016)
15. Zhou, X., Ito, T., Takayama, R., Wang, S., Hara, T., Fujita, H.: Three-dimensional CT image segmentation by combining 2D fully convolutional network with 3D majority voting. In: Carneiro, G., et al. (eds.) LABELS/DLMIA -2016. LNCS, vol. 10008, pp. 111–120. Springer, Cham (2016). https://doi.org/10.1007/978-3-319-46976-8_12