Received: 17 July 2023 | Revised: 5 December 2023 | Accepted: 4 February 2024

DOI: 10.1002/mp.17000

RESEARCH ARTICLE

MEDICAL PHYSICS

Deep-learning-based joint rigid and deformable contour propagation for magnetic resonance imaging-guided prostate radiotherapy

Iris D. Kolenbrander[1,2] | **Matteo Maspero**[3,4] | **Allard A. Hendriksen**[5] | **Ryan Pollitt**[1] | **Jochem R. N. van der Voort van Zyp**[4] | **Cornelis A. T. van den Berg**[3,4] | **Josien P. W. Pluim**[1,2] | **Maureen A. J. M. van Eijnatten**[1,2]

[1]Medical Image Analysis Group, Department of Biomedical Engineering, Eindhoven University of Technology, Eindhoven, The Netherlands

[2]Eindhoven Artificial Intelligence Systems Institute, Eindhoven University of Technology, Eindhoven, The Netherlands

[3]Computational Imaging Group for MR Diagnostics & Therapy, Center for Image Sciences, University Medical Center Utrecht, Utrecht, The Netherlands

[4]Department of Radiotherapy, University Medical Center Utrecht, Utrecht, The Netherlands

[5]Computational Imaging, Centrum Wiskunde & Informatica, Amsterdam, The Netherlands

Correspondence
Iris D. Kolenbrander, Medical Image Analysis Group, Eindhoven University of Technology, Eindhoven, The Netherlands.
Email: i.d.kolenbrander@tue.nl

Funding information
The Irene Curie Fellowship program; Eindhoven Artificial Intelligence Systems Institute; Nederlandse Organisatie voor Wetenschappelijk Onderzoek, Grant/Award Number: 639.073.506

Abstract

Background: Deep learning-based unsupervised image registration has recently been proposed, promising fast registration. However, it has yet to be adopted in the online adaptive magnetic resonance imaging-guided radiotherapy (MRgRT) workflow.

Purpose: In this paper, we design an unsupervised, joint rigid, and deformable registration framework for contour propagation in MRgRT of prostate cancer.

Methods: Three-dimensional pelvic T2-weighted MRIs of 143 prostate cancer patients undergoing radiotherapy were collected and divided into 110, 13, and 20 patients for training, validation, and testing. We designed a framework using convolutional neural networks (CNNs) for rigid and deformable registration. We selected the deformable registration network architecture among U-Net, MS-D Net, and LapIRN and optimized the training strategy (end-to-end vs. sequential). The framework was compared against an iterative baseline registration. We evaluated registration accuracy (the Dice and Hausdorff distance of the prostate and bladder contours), structural similarity index, and folding percentage to compare the methods. We also evaluated the framework's robustness to rigid and elastic deformations and bias field perturbations.

Results: The end-to-end trained framework comprising LapIRN for the deformable component achieved the best median (interquartile range) prostate and bladder Dice of 0.89 (0.85–0.91) and 0.86 (0.80–0.91), respectively. This accuracy was comparable to the iterative baseline registration: prostate and bladder Dice of 0.91 (0.88–0.93) and 0.86 (0.80–0.92). The best models complete rigid and deformable registration in 0.002 (0.0005) and 0.74 (0.43) s (Nvidia Tesla V100-PCIe 32 GB GPU), respectively. We found that the models are robust to translations up to 52 mm, rotations up to 15°, elastic deformations up to 40 mm, and bias fields.

Conclusions: Our proposed unsupervised, deep learning-based registration framework can perform rigid and deformable registration in less than a second with contour propagation accuracy comparable with iterative registration.

KEYWORDS
contour propagation, deep-learning-based image registration, MRI-guided radiotherapy

1 | INTRODUCTION

Prostate cancer is the second most commonly diagnosed cancer and the fifth most common cause of cancer death in men worldwide.[1] One of the standard treatments for prostate cancer is external-beam radiotherapy (RT), which involves delivering radiation over multiple sessions (fractions).[2] In the last decade, hypofractionation, that is, using few irradiations, has been increasingly adopted.[3] For prostate RT, the noninferiority of hypofractionation has been proven, allowing the number of fractions from conventional (35 fractions) to moderate (20 fractions) and ultra-hypofractionated (five fractions) treatments.[4]

Ultra-hypofractionated treatments require accurate target localization, made possible by integrating imaging and treatment devices. For example, magnetic resonance imaging-guided RT (MRgRT) has recently become viable, providing superb soft tissue contrast during treatment. Such devices enable online adaptive RT, which involves adapting the treatment plan to the daily anatomy to reduce the dose to healthy surrounding tissue.[5,6] The existing MRgRT systems can adapt the treatment plan to the current anatomy (just before treatment) through daily imaging and time-consuming manual recontouring (10 min) of the target and organs at risk.[7]

Current research focuses on automating the workflows to adapt the treatment plan with minimal manual intervention. In this sense, rapid and accurate automated contouring solutions should be developed. Deformable image registration accomplishes this objective by propagating the contours from the simulation (offline) MRI to the daily (online) MRI. Speed is essential to reduce treatment time and motion between the scan and the irradiation. For example, Smith et al.[8] found that the bladder fills up to 74% of the original volume between the daily scan and irradiation. Current online adaptive MRgRT workflows use slow iterative registration (in the order of minutes).[9] However, deep learning-based registration has recently become popular for time-constrained applications, for example, image guidance in interventional settings and motion correction in MRgRT.[10–12]

Deep learning-based registration methods have primarily concentrated on deformable registration, performing rigid pre-alignment using iterative registration, which slows down the overall registration.[13–15]

Furthermore, which network architecture is optimal for contour propagation in online adaptive MRgRT needs to be clarified. Many existing registration approaches rely on encoder–decoder network architectures such as the U-Net.[13,14,16,17] Other promising network architectures have yet to be thoroughly investigated for prostate MRgRT.[18,19] The registration method should accurately propagate the contours in this setting, requiring minimal manual adjustments. The accuracy should also be maintained in various scenarios to ensure safe clinical implementation, which requires robust models that generalize well to diverse data distributions, for example, different sources and magnitudes of motion.

This paper investigates convolutional neural networks (CNNs) for unsupervised, joint rigid, and deformable image registration to facilitate accurate contour propagation in prostate MRgRT. We design a framework consisting of two (rigid and deformable) CNNs, for which we optimize the training strategy by comparing sequential training of the CNNs with end-to-end training. Our main contribution is a thorough comparison of three promising CNN architectures for deformable registration within our framework in terms of registration accuracy and speed. We also performed robustness experiments to investigate the applicability of the networks in real world scenarios.

2 | METHODS

2.1 | Data and preprocessing

Data collection

A dataset of 143 sequential patients who underwent radiotherapy for prostate cancer at the University Medical Center Utrecht, The Netherlands, between January 2020 and May 2021, was collected. These patients were part of the ethics review board-approved MOMENTUM (Multiple Outcome Evaluation of Radiotherapy Using the MR-Linac) study.[20] They underwent 5 × 7.25-Gy fractions on a Unity 1.5 T MR-Linac system (Elekta AB, Stockholm, Sweden) using an "adapt-to-shape" workflow.[21]

In the planning phase, patients underwent MRI on a 3 T scanner (Ingenia, Philips Medical Systems, Best, The Netherlands) to delineate the target and organs at risk (rectum, bladder, and femurs). In each fraction on the MR-Linac, an initial daily MRI was acquired for recontouring the patient and adapting RT planning to the daily anatomy, followed by a second MRI before irradiation to ensure accurate patient positioning. The target and OAR contours were propagated to the initial daily MRI using the clinical treatment planning system (Monaco, Elekta AB, Stockholm, Sweden). The radiation oncologist manually adjusted the propagated contours within a ring defined as the isotropic extension of 2 cm of the target, providing an accurate label only for the prostate.

Prostate dataset

For each patient, the planning MRI and five initial daily MRIs were included, resulting in 858 T2-weighted 3D turbo spin-echo MRIs and their prostate labels. We

TABLE 1 Image acquisition parameters of the planning MRI, and the initial (init) and position verification (PV) daily fraction MRIs.

Parameter		Planning MRI	Daily fraction MRI	
			Init	PV
Patients (n)		143	143	10
Images (n)		143	715	50
Magnetic field strength (T)		3	1.5	
Relaxation time (ms)		1700–2500	1535–1635	
Echo time (ms)		247–360	120–277.8	
Flip angle (°)		90	90	
Bandwidth (Hz/px)		384–740	562–740	
Field of view[a] (mm^3)	Lower	$400 \times 400 \times 200$	$400 \times 400 \times 180$	
	Upper	$451 \times 451 \times 300$	$448 \times 448 \times 300$	
Voxel spacing[a] (mm^3)	Lower	$0.59 \times 0.59 \times 2$	$0.75 \times 0.75 \times 2$	
	Upper	$0.83 \times 0.83 \times 2$	$0.83 \times 0.83 \times 2$	
Scan time (s)		190–294	116–204	

[a]Expressed in mediolateral, anteroposterior, and craniocaudal directions.

randomly divided the dataset into 110, 13, and 20 patients for training, validation, and testing, corresponding to 550, 65, and 100 image pairs. The acquisition parameters varied (Table 1) due to protocol modification, and the variations' distribution was verified after the split. All MRIs were resampled to isotropic voxel sizes of $1 \times 1 \times 1$ mm, center-cropped to a FOV of $208 \times 208 \times 208$ mm, corrected for bias field,[22] intensity clipped to the 95th percentile and normalized to the range [0–1].

Bladder subset

For 10 patients in the test set, manual bladder delineations were available on the daily position verification (PV) MRI of each fraction as part of the prospective Utrecht Prostate Cohort Study (NCT04228211).[23] The imaging parameters of the PV MRI were identical to the daily MRI (Table 1).

Data augmentation

On-the-fly data augmentation was performed to increase the variation of the training set. We generated three additional image pairs for each image pair by augmenting either the fixed image, the moving image, or both, resulting in 2200 (550×4) image pairs for one training epoch. Each augmentation involved deforming the image with a random deformation and adding Gaussian noise (sigma = 0.005). The deformation was defined on a 5^3 control point grid of displacements sampled from a Gaussian distribution, which were rescaled to a maximum of 20 mm and upsampled using third-order B-spline interpolation.

2.2 | Deep-learning-based registration

2.2.1 | Joint rigid and deformable registration framework

The framework consists of a cascade of two CNNs for rigid and deformable image registration (Figure 1). The CNNs predict the rigid transformation (T_R) and a dense displacement vector field (DVF: ϕ), which propagate the contours of the planning MRI (moving: M) to the daily MRI (fixed: F).

Rigid component

The rigid component consists of a CNN for rigid transformation prediction and a spatial transformer module (STM) that obtains the rigidly aligned moving image through interpolation.[24] The rigid CNN is an encoder consisting of four convolutional layers with kernel size 3^3, dilation rate 2, LeakyReLU activation (slope = −0.2), and a fully connected layer (FCL). It predicts the translation and rotation vectors, forming the rigid transformation matrix (Figure 2).[25]

We tuned the encoder's width (2–32), number of FCLs (1–2), and the location of concatenating the fixed and moving features (first layer and layer before the FCL), finding an optimum width of 8, one FCL, and concatenation of the image features in the first layer. In this tuning phase, the encoder was trained with random rigid transformations, using the error between the predicted and ground truth transformations as the objective function and evaluation metric for validation. The rigid transformations contained translations up to 31.2 mm and rotations up to 15°.

Deformable component

The deformable component consists of a CNN that takes the fixed and the rigidly aligned moving image and predicts the DVF, followed by an STM that deforms the moving image. In the STM, the rigid transformation matrix and DVF were combined to reduce the potential accumulation of interpolation artifacts.

Most deformable registration methods for prostate MRgRT rely on U-Net,[13,14,16,17] but there are limited comparisons to other CNN architectures. We investigated three promising 3D architectures for deformable registration: a U-Net, a mixed-scale dense (MS-D) network, and a Laplacian pyramid image registration network (LapIRN). The U-Net is considered a baseline. The MS-D network is of interest for the speed required for this task and its overall good performance while having fewer trainable parameters (about 5% of U-Net's parameters).[19] The LapIRN[18] is considered one of the best-performing deformable networks, given the outstanding results obtained within the recent Learn2Reg Challenge.[26]

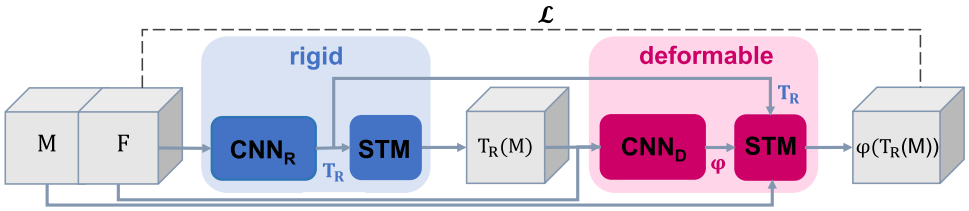

FIGURE 1 The joint rigid and deformable registration framework consisting of a CNN for the rigid registration (CNN_R) and a CNN for the deformable registration (CNN_D) that predict the rigid transformation matrix (T_R) and the DVF (ϕ). The spatial transformer modules (STMs) obtain the rigidly transformed moving image ($T_R(M)$) and the final deformed moving image ($\phi(T_R(M))$) through image interpolation.

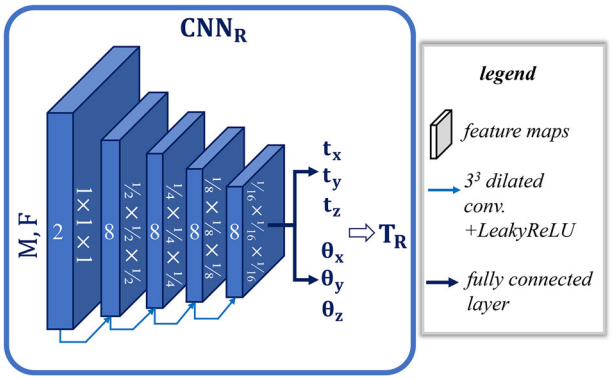

FIGURE 2 The rigid CNN architecture (CNN_R) is used within the joint rigid and deformable registration framework. CNN_R takes the fixed (F) and moving (M) images and predicts the rigid transformation matrix (T_R).

U-Net

We used VoxelMorph's implementation of U-Net, which consists of an encoder and decoder that apply convolutions with kernel size 3^3, followed by a LeakyReLU activation (slope = −0.2), and skip connections in between[27] (Figure 3). The encoder consists of four convolutional layers with a stride of 2, halving the spatial dimensions at each layer. The decoder consists of four successive upsampling and convolutional layers, obtaining full-resolution feature maps and three additional convolutional layers.

MS-D Net

We adapted the original 2D implementation to perform 3D deformable registration.[28] Each feature map takes all previous features and the input layer and convolves them using dilated convolutions with kernel size 3^3, dilation rate r, and rectified linear unit (ReLU) activation (Figure 3). The dilation rate in layer i equals $r = (i\%10) + 1$. The feature maps are densely connected to the output layer and are equal in size due to reflective padding and the absence of pooling layers. The network depth was tuned using depths ranging from 20 to 150 on a subset of 110 image pairs of the prostate dataset, finding the optimal depth of 30 based on the prostate Dice and speed (Section 2.4). In the tuning phase, the models were trained with the objective function presented in Section 2.2.2 The network width of 1 was based on the work of Minnema et al.[29] and allowed training in a reasonable time, for example, 1–2 weeks.

LapIRN

LapIRN contains three levels of CNNs operating at different resolutions to mimic conventional multiresolution registration (Figure 3). Using the original implementation,[18] we parameterized the model with displacement fields. We used a sequential training scheme (using the objective function presented in Section 2.2.2), adding a level after training the previous level for 30 000 iterations. The parameters of the previous levels were frozen for the subsequent 2000 iterations.

Each level consists of an encoder, five residual blocks (RB), and a decoder and applies convolutions with kernel size 3^3 and LeakyReLU activation (slope = −0.2). The encoder consists of two convolutional layers with stride 1, followed by a convolutional layer with stride 2. The residual block contains two convolutional layers with pre-activation and a skip connection. Finally, the decoder uses a transposed convolutional layer and two successive convolutional layers with stride 1, followed by a SoftSign activation multiplied by a factor of 10.

2.2.2 | Unsupervised training

The framework is trained using unsupervised learning, optimizing the similarity between the fixed (daily MRI) and deformed moving images (registered planning MRI). We optimized the training strategy by comparing end-to-end training of the rigid and deformable components (after random initialization) or sequential training (including a warm-up of the rigid component) in Experiment 2.5.1.

Objective function

The framework optimizes the objective function ($\mathcal{L}$ in Equation 1), including an image similarity term, namely the normalized mutual information (NMI) between the fixed image and the deformed moving image ($\phi(T_R(M))$), where $T_R(M)$ is the rigidly aligned moving image, and a regularization term ($\mathcal{L}_{reg}$), namely the L2-norm of the

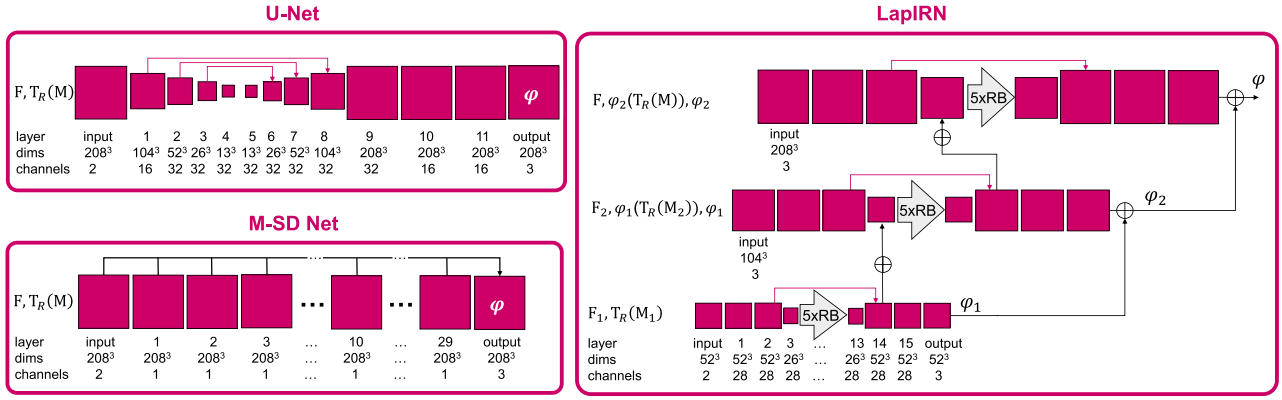

FIGURE 3 The CNN architectures used for the deformable component within the registration framework. All networks take the fixed image (F) and rigidly aligned moving image ($T_R(M)$) as input and predict a dense DVF (ϕ). The U-Net has a typical encoder–decoder structure, whereas the MS-D Net is characterized by equally sized feature maps and dense connections. The LapIRN contains three levels that operate at different resolutions.

DVF gradients scaled by a factor λ.

$$\mathcal{L} = NMI(F, \phi(T_R(M))) + \lambda \mathcal{L}_{reg}(\phi) \quad (1)$$

To prevent overfitting, we stopped the training when the change in prostate Dice (Section 2.4) running average was negative for four consecutive epochs.

Training details

The framework was trained using the Adam optimizer and a batch size of 1 on a PC with an Nvidia Tesla V100-PCIe GPU. We tuned the learning rate and the regularization weight (λ) for all models in eight training runs via a random search on a range of [1e-4; 1e-3] and [0.5; 10], respectively. To do this, we trained the models in 50 epochs on a subset of the prostate dataset, using 110 image pairs instead of 550 to reduce the training times. A learning rate of 1e-4 was optimal for all models. Regularization weights 1, 1, and 5 were optimal for the MS-D Net, LapIRN, and U-Net.

2.3 | Iterative baseline registration

We used rigid registration followed by deformable registration performed with iterative registration as a baseline. We considered the open-source Advanced Normalization Tools (ANTs)[30,31] and Elastix[32] as software packages. We found similar contour propagation accuracies, with slightly better performance for Elastix (Supplementary materials; Figure S1). We selected Elastix as the iterative baseline. The registration was performed using multithreading on a system with 257 GB RAM and 72 CPU cores. NMI was used as the objective function, and a multiresolution strategy was applied with four resolutions and 2000 iterations, with B-splines on a final grid spacing of 8 mm.[9]

2.4 | Evaluation metrics

The prostate and bladder contour propagation accuracy was evaluated using the Dice similarity coefficient (Dice) of the fixed and propagated segmentations and the 95th percentile Hausdorff distance (95%HD) of the fixed and propagated contours. In addition, the registration quality was evaluated using the structural similarity index measure (SSIM) and the percentage negative values of the Jacobian determinant of the DVF ($\%|J| < 0$), which is indicative of undesired tissue folding. The average training time per epoch (2200 iterations) was computed and the registration speed was measured as the mean (standard deviation; std) latency on a GPU (Nvidia Tesla V100-PCIe 32GB) for all image pairs in the prostate test set. We performed statistical analysis using Friedman's test followed by post hoc pairwise comparisons with the paired Bonferroni corrected Wilcoxon signed-rank tests when Friedman's test was significant.

2.5 | Experiments

2.5.1 | End-to-end versus sequential training

We trained the framework with each of the three CNNs for the deformable component for 300 epochs on a subset of the prostate dataset containing 110 image pairs to optimize the training strategy. We evaluated the registration accuracy on the validation set. The rigid and deformable components of the framework were trained end-to-end (after random initialization) or sequentially using the loss function presented in Equation (1).

Sequential training included a warm-up phase for the rigid component during which the deformable

component was frozen (first 40 epochs), followed by jointly training the rigid and deformable components (subsequent 260 epochs). We considered three variants of sequential training: freezing the rigid component in the joint phase (S1), freezing the rigid component in only the first 150 epochs of the joint phase (S2), or no freezing of the rigid component (S3). Note that the training strategy only affects the number of trainable parameters depending on the training phase.

2.5.2 | Deformable registration: Network architecture comparison

We compared three CNNs for the deformable registration within the joint rigid and deformable registration framework: U-Net, MS-D Net, and LapIRN (Section 2.2.1). The framework containing each of these CNNs was trained on the prostate training set, and their hyperparameters were optimized on the prostate validation set, and their registration accuracy and speed were evaluated on the prostate and bladder test sets using iterative registration as a baseline (Section 2.4).

2.5.3 | Robustness to input perturbations

We investigated the robustness of the rigid and deformable registration. For the rigid registration, we studied the range of rigid transformations that could be recovered using the best-performing framework resulting from the previous experiments 2.5.1 and 2.5.2. We used the original test image pairs rigidly aligned by the rigid component as baselines, such that they contained minimal rigid misalignments. We then applied a random rigid transformation to the fixed image: either a rotation (5–20°) or a sagittal-plane translation (10.4–104 mm), given that translations were most likely to occur in this plane. The applied translations and rotations were taken as ground-truth. Finally, the rigid component registered the moving image to the perturbed fixed image, and we evaluated the mean squared error (MSE) between the ground truth and predicted translation or rotation vectors. We also evaluated the Dice between the perturbed fixed and the registered moving prostate contour to identify the conditions for achieving acceptable contour accuracy.

We also evaluated the robustness of the deformable component (all CNNs; Section 2.2.1) to nonlinear deformations or synthetic bias fields. The rigidly aligned test image pairs (using the rigid component) served as a baseline. From each pair, the moving image was perturbed and then registered to the original fixed image by the deformable component, after which the prostate Dice and percentage foldings were evaluated. The nonlinear deformations were defined on a grid of displacements sampled from a Gaussian distribution, rescaled to a maximum displacement, and upsampled using third-order B-spline interpolation. The deformations were varied: the number of grid points ranged from 5^3 (coarse-scale deformation) to 25^3 (fine-scale deformation) while using a maximum displacement of 20 mm, or the maximum displacement ranged from 5 to 60 mm on a 5^3 grid. The bias field artifacts were modeled as linear combinations of third-order polynomial basis functions and varied in magnitude: the polynomial coefficients ranged from 0.2 to 0.7.[33]

Estimating the training set distribution

To gain insight into the framework's robustness within and beyond the training set distribution, we estimated the range of rigid and nonlinear motions available during training. We estimated the upper bound of the rigid transformations in the training set as the 95th percentile of the translations and rotations (vector norms) predicted by Elastix, which were found to be 78.3 mm and 3.54° (Supplementary materials; Table S1). We considered the nonlinear deformations of the data augmentation procedure to be the training range (5^3 grid; max. displacement of 20 mm).

3 | RESULTS

3.1 | End-to-end versus sequential training

Sequential and end-to-end training produced comparable prostate Dice values (Tables 2 and S2). We performed the subsequent experiments using the end-to-end training strategy because it does not require freezing and unfreezing of the components.

3.2 | Deformable registration: Network architecture comparison

LapIRN achieved the highest prostate and bladder Dice values ($p < 0.01$), the lowest 95%HDs ($p < 0.01$), and the highest SSIM ($p < 0.01$) of all CNNs within the framework (Figure 4). The resulting framework achieved a median (IQR) Dice of 0.89 (0.85–0.91) for the prostate and 0.86 (0.80–0.91) for the bladder, which were not significantly different from the iterative baseline prostate and bladder Dice values: 0.91 (0.88–0.93) and 0.86 (0.80–0.92).

Figure 5 shows a typical example of the propagation of a prostate contour, for which all models agreed with the ground truth manual delineation. LapIRN produced the least smooth DVF, with most foldings near the bladder. The framework infers within 1 s per volume, substantially faster than iterative registration (Table 3).

TABLE 2 Effect of end-to-end and sequential training configurations on the prostate Dice (↑). Values are median (IQR).

Training config.	U-Net	MS-D	LapIRN	
End-to-end		0.87 (0.81–0.90)	0.85 (0.77–0.88)	0.88 (0.84–0.90)
Seq.	**S1**	0.88 (0.81–0.91)	0.85 (0.75–0.87)*	0.88 (0.83–0.90)
	S2	0.88 (0.83–0.90)	0.85 (0.77–0.87)	0.88 (0.84–0.90)
	S3	0.87 (0.83–0.90)	0.84 (0.79–0.87)	0.89 (0.84–0.91)

*Significant differences from end-to-end configuration are shown ($p < 0.05$).

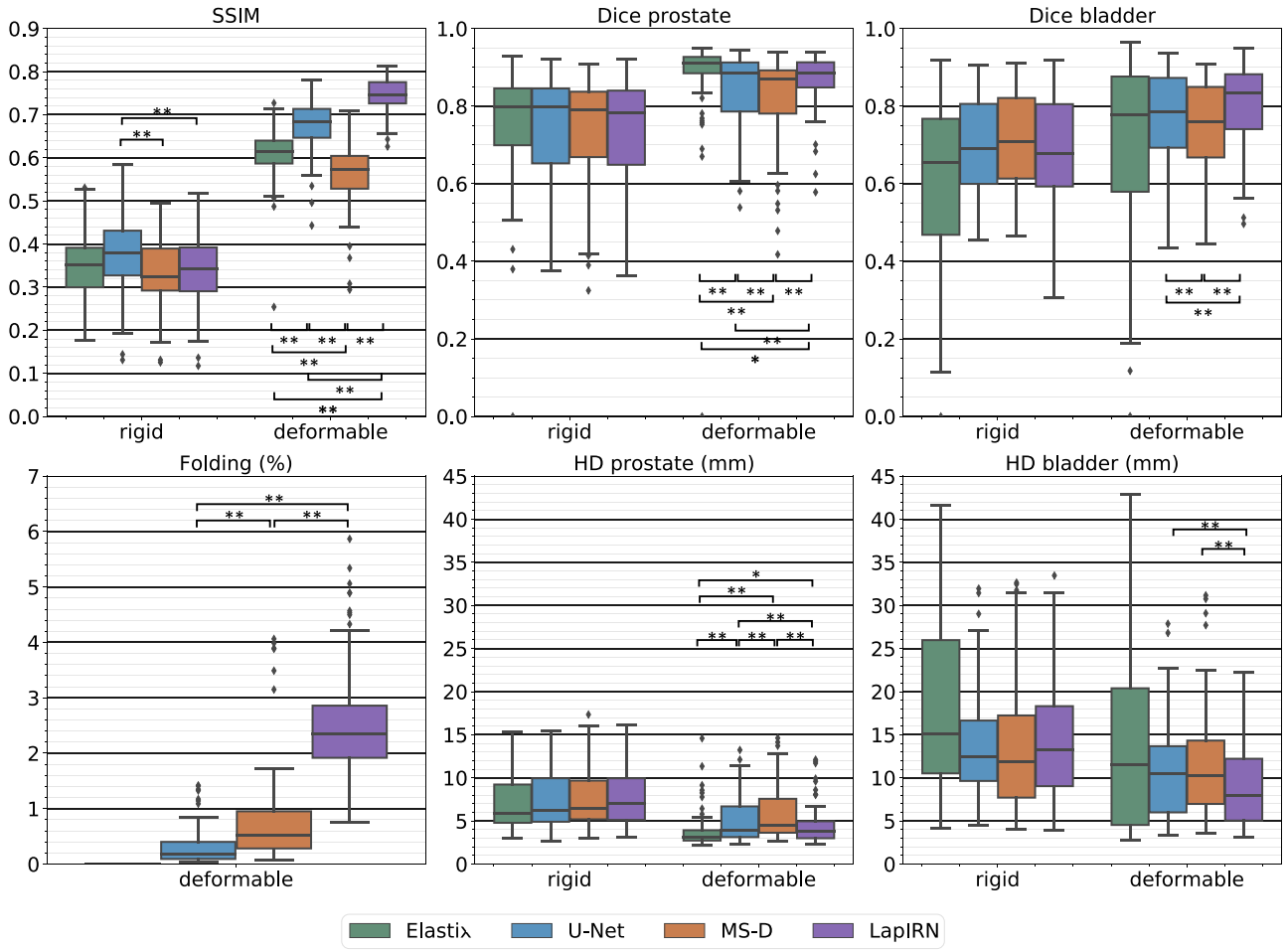

FIGURE 4 Quantitative comparisons of the three CNNs for deformable registration within the joint rigid and deformable registration framework and iterative registration. The boxplots show the SSIM and percentage foldings, the prostate Dice, and 95%HD evaluated on the prostate test set (100 image pairs). Bladder Dice and 95%HD are evaluated on the bladder test set (50 image pairs). *$p < 0.05$; **$p < 0.01$.

TABLE 3 The mean (std) latency, training times, and the number of trainable parameters. All DL-based training times and latencies are measured on a PC with an Nvidia Tesla V100-PCIe 32 GB GPU.

			Training time per epoch	Training epochs	Latency (s)	Trainable parameters
DL-based	Rigid		n.a.	n.a.	0.002 (0.0005)	111 110
	Deformable	LapIRN	3 h 20 min	40	0.74 (0.43)	615 076
		U-Net	2 h 30 min	42	0.34 (0.21)	301 411
		MS-D Net	2 h 45 min	60	0.22 (0.03)	13 494
Elastix	Rigid		n.a.	n.a.	28.0 (0.8)*	n.a.
	Deformable		n.a.	n.a.	34.0 (0.6)*	n.a.

*Measured on a PC with a CPU equipped with 257 GB RAM and 72 CPU cores.

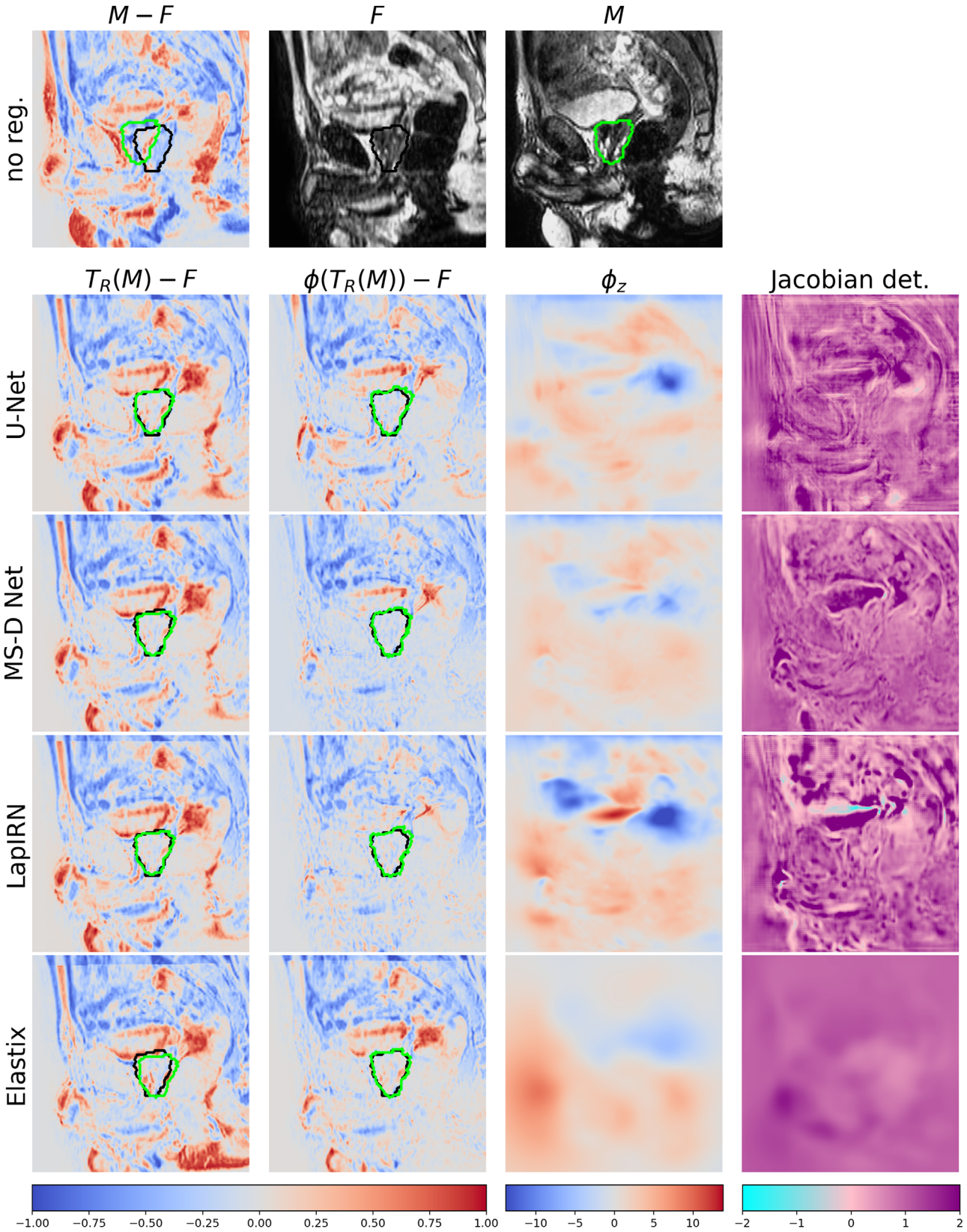

FIGURE 5 Registration of a prostate test case (sagittal slice). The difference images after rigid (first column) and deformable (second column) registration, with the prostate contours in black (fixed) and green (moving; deformed). The craniocaudal displacements (ϕ_z) and the Jacobian determinant of the DVF are shown in the third and fourth columns.

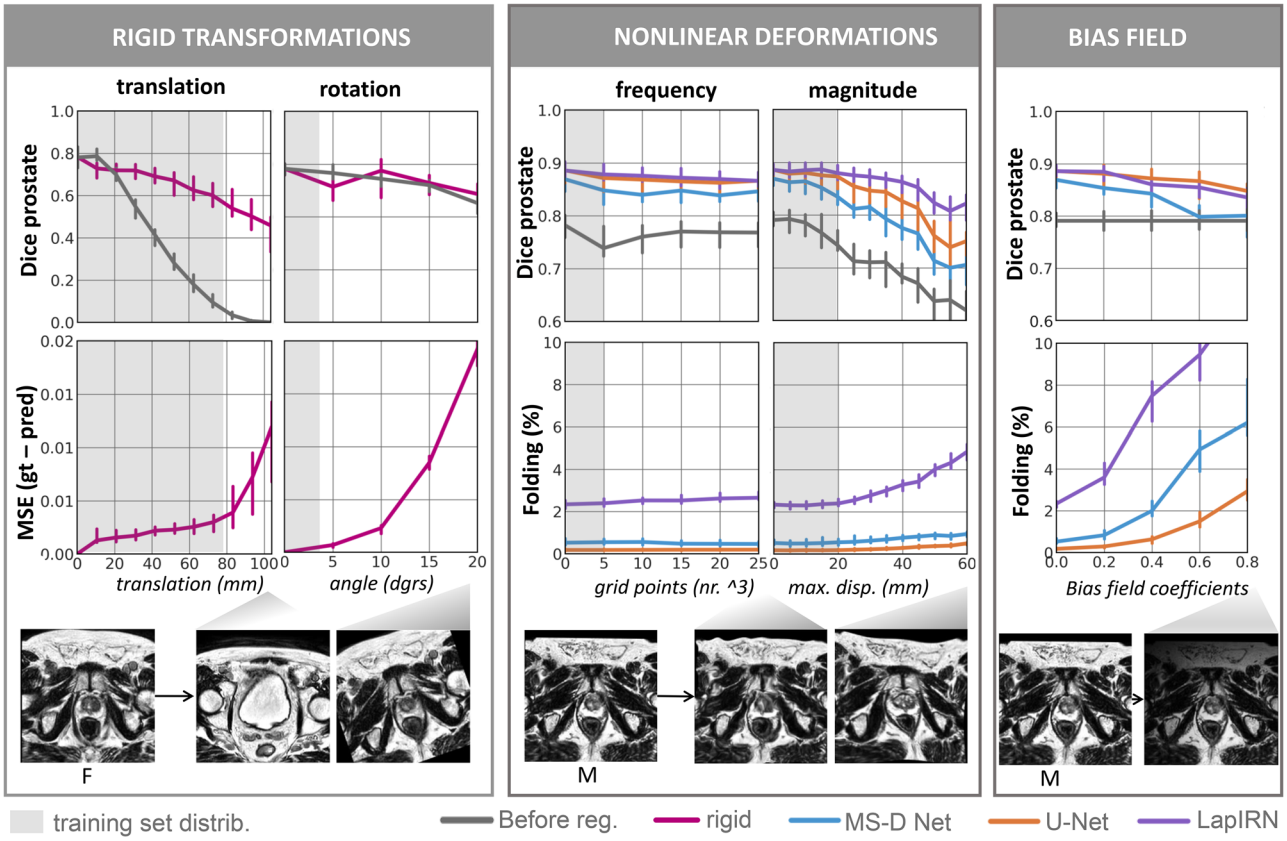

FIGURE 6 The models' robustness to rigid transformations, nonlinear deformations, and bias field perturbations. The lines (error bars) show the median (95% confidence intervals) of the prostate Dice, percentage foldings, and the mean squared error (MSE) of the ground truth (gt) and predicted (pred) translations/rotations. The semitransparent boxes show the motion ranges within the training set. The bottom row shows examples (axial slices) of the perturbed fixed (F) or moving (M) image.

3.3 | Robustness to input perturbations

The rigid registration (of the best-performing framework with LapIRN) recovered the translations and rotations seen during training, for example, the median (IQR) prostate Dice decreased from 0.78 (0.65–0.84) at 0 mm translations to 0.76 (0.51–0.77) at 52 mm translations (Figure 6). For the deformable registration, most CNNs were robust to the nonlinear deformations in the training set, for example, U-Net and LapIRN were unaffected by deformations under 20 mm.

Transformation and deformation magnitudes beyond the range covered in the training set affected most models, for example, the prostate Dice of the rigid registration decreased to 0.73 (0.60–0.81) at 15° rotations (Figure 6). LapIRN is the most robust among the CNNs for the deformable registration, achieving a prostate Dice of 0.86 (0.81–0.89) for 40 mm nonlinear deformations and 0.86 (0.81–0.90) for bias fields with a coefficient of 0.4. However, LapIRN produced a factor of 3.2 more foldings when bias fields (coefficient 0.4) were introduced.

4 | DISCUSSION

This work proposes an unsupervised, joint rigid, and deformable image registration framework for contour propagation in prostate MRgRT. We optimized the framework's training strategy and evaluated three promising network architectures for deformable registration. Overall, the framework incorporating the LapIRN network for the deformable component resulted in the highest accuracy and robustness to (simulated) nonlinear deformations and bias field artifacts. Using a coarse-to-fine cascaded approach may be the source of this result since other work has found similar advantages of such approaches over single networks.[18,25]

We also demonstrated that end-to-end training and sequential training of the rigid and deformable components of the framework lead to similar registration accuracies. Other works have trained the individual components of their proposed frameworks sequentially (similar to the configurations S1–3 in Table 2) to avoid exploding gradients.[25,34] In our work, we did not encounter any exploding gradients. Note that this discrepancy could be related to different networks

employed. We recommend end-to-end training of the rigid and deformable components, as this achieves similar performance and avoids component (un)freezing.

MRgRT requires accurate contour delineation. Our unsupervised registration framework achieves prostate contour accuracies (Dice 0.89) comparable to those reported by others (Dice≈0.87).[13,14,16] The prostate accuracy may be considered clinically acceptable compared to the interobserver agreement of manual delineations (reported interobserver Dices range from 0.87 to 0.94[35–37]). In addition, Thornqvist et al.[38] reported that a mean prostate Dice of 0.86 is considered acceptable by a radiation oncologist.

The obtained bladder contours (Dice 0.86) still require manual adjustments, as suggested by a mean (std) interobserver Dice of 0.93 (0.03) found by Roach et al.[36] Bladder filling/emptying causes large deformations that challenge deep learning-based registration[39] and may explain why LapIRN produced significant folding near the bladder. Image registration methods often assume that the topology of structures is preserved and consequently aim to find smooth (continuous and invertible) deformation fields without folding.[40] This approach ensures physically plausible deformations but may face challenges for organs undergoing substantial volume changes, as seen in cases like bladder filling. The bladder contour accuracy achieved by the framework was comparable to that of iterative registration. The iterative baseline registration was optimized to propagate the prostate, which may have caused this nonoptimal bladder propagation. We emphasize the importance of drinking protocols to minimize the scan differences caused by bladder filling/emptying.

Currently, contour adaptation takes about 10 min in the MRgRT workflow.[7] Our deep learning framework achieves sub-second contour propagation with an accuracy that is on par with iterative registration. The bladder contours should be improved before clinical use to minimize manual adaptations, which counteract the theoretic speed gain of automated contouring. For example, weak supervision with the contours could achieve this improvement.[14] Combining the advantages of registration and segmentation to obtain fused models could also be powerful.[41]

We identified the conditions under which the framework achieves "acceptable" contour accuracy—defined as a minimum prostate Dice of 0.67 (25th percentile of the iterative rigid registration) for the rigid registration and 0.86 (considered acceptable by a radiation oncologist[38]) for the deformable registration. The framework comprising LapIRN achieved acceptable contour accuracies for translations up to 52 mm, rotations up to 15°, nonlinear deformations up to 40 mm, and bias fields to some extent. We emphasize the importance of robustness experiments to understand better the capabilities and limitations of deep learning models in real-world scenarios. Only two other papers have evaluated the robustness of their contour propagation method to different motion magnitudes: Eppenhof et al. found their method to be robust to (simulated) craniocaudal prostate shifts,[14] and Elmahady et al.[13] found their method to be robust to bladder filling. In our in silico robustness experiments, we found our framework robust to transformations in the training set, which reflect clinically occurring motion.

5 | CONCLUSION

In this work, we designed an unsupervised, joint rigid, and deformable registration framework based on deep learning that achieves sub-second contour propagation. The framework propagates the prostate contours with good accuracy and the bladder contours with an accuracy comparable to iterative registration. The framework can propagate the prostate contours with acceptable accuracy for translations up to 52 mm, rotations up to 15°, nonlinear deformations up to 40 mm, and bias fields. Deep learning enables fast and accurate contour propagation, which is critical in online adaptive MRgRT to reduce daily treatment times and improve conformity to the daily anatomy.

ACKNOWLEDGMENTS

The Irene Curie Fellowship program and the Eindhoven Artificial Intelligence Systems Institute partially funded this work. M. A. J. M. E acknowledges financial support from the Netherlands Organisation for Scientific Research (NWO), project number 639.073.506.

CONFLICT OF INTEREST STATEMENT

The authors declare no conflicts of interest.

REFERENCES

1. Rawla P. Epidemiology of prostate cancer. *World J Oncol*. 2019;10:63-89.
2. Litwin MS, Tan H. The diagnosis and treatment of prostate cancer: a review. *J Am Med Assoc*. 2017;317:2532-2542.
3. Jackson WC, Silva J, Hartman HE, et al. Stereotactic body radiation therapy for localized prostate cancer: a systematic review and meta-analysis of over 6,000 patients treated on prospective studies. *Int J Radiat Oncol Biol Phys*. 2019;104:778-789.
4. Dearnaley D, Syndikus I, Mossop H, et al. Conventional versus hypofractionated high-dose intensity-modulated radiotherapy for prostate cancer: 5-year outcomes of the randomised, non-inferiority, phase 3 CHHiP trial. *Lancet Oncol*. 2016;17:1047-1060.
5. Raaymakers BW, Lagendijk JJW, Overweg J, et al. Integrating a 1.5 T MRI scanner with a 6 MV accelerator: proof of concept. *Phys Med Biol*. 2009;54:N229-N237.
6. Raaymakers Bas W, Jürgenliemk-Schulz IM, Bol GH, et al. First patients treated with a 1.5 T MRI-Linac: clinical proof of concept of a high-precision, high-field MRI guided radiotherapy treatment. *Phys Med Biol*. 2017;62:L41-L50.
7. Güngör G, Serbez İ, Temur B, et al. Time analysis of online adaptive magnetic resonance-guided radiation therapy workflow according to anatomical sites. *Pract Radiat Oncol*. 2021;11:e11-e21.

8. Smith JA, Dunlop A, Barnes H, et al. Bladder filling in patients undergoing prostate radiotherapy on a MR-linac: the dosimetric impact. *Technical Innov Patient Support Radiat Oncol*. 2022;21:41-45.
9. Klein S, Van Der Heide UA, Lips IM, Van Vulpen M, Staring M, Pluim JPW. Automatic segmentation of the prostate in 3D MR images by atlas matching using localized mutual information. *Med Phys*. 2008;35:1407-1417.
10. Terpstra ML, Maspero M, Bruijnen T, Verhoeff JJC, Lagendijk JJW, Berg CAT. Real-time 3D motion estimation from undersampled MRI using multi-resolution neural networks. *Med Phys*. 2021;48:6597-6613.
11. Chen S, Eldeniz C, Fraum TJ, et al. Respiratory motion management using a single rapid MRI scan for a 0.35 T MRI-Linac system. *Med Phys*. 2023;50:6163-6176.
12. Hall WA, Paulson ES, van der Heide UA, et al. The transformation of radiation oncology using real-time magnetic resonance guidance: a review. *Eur J Cancer*. 2019;122:42-52.
13. Elmahdy MS, Jagt T, Zinkstok RTh, et al. Robust contour propagation using deep learning and image registration for online adaptive proton therapy of prostate cancer. *Med Phys*. 2019;46:3329-3343.
14. Eppenhof KAJ, Maspero M, Savenije MHF, et al. Fast contour propagation for MR-guided prostate radiotherapy using convolutional neural networks. *Med Phys*. 2020;47:1238-1248.
15. Chen J, Frey EC, He Y, Segars WP, Li Y, Du Y. TransMorph: transformer for unsupervised medical image registration. *Med Image Anal*. 2022;82:102615.
16. Kawula M, Vagni M, Cusumano D, et al. Prior knowledge based deep learning auto-segmentation in magnetic resonance imaging-guided radiotherapy of prostate cancer. *Phys Imaging Radiat Oncol*. 2023;28:100498.
17. Hemon C, Rigaud B, Barateau A, et al. Contour-guided deep learning based deformable image registration for dose monitoring during CBCT-guided radiotherapy of prostate cancer. *J Appl Clin Med Phys*. 2023;24:e13991.
18. Mok TCW, Chung A. Large deformation diffeomorphic image registration with Laplacian pyramid networks. In: International Conference on Medical Image Computing and Computer-Assisted Intervention (MICCAI). Springer; 2020:211-221.
19. Pelt DM, Sethian JA. A mixed-scale dense convolutional neural network for image analysis. *Proc Natl Acad Sci*. 2018;115:254-259.
20. Otterloo SR, Christodouleas JP, Blezer ELA, et al. The MOMENTUM study: an international registry for the evidence-based introduction of MR-guided adaptive therapy. *Front Oncol*. 2020;10:1328.
21. Winkel D, Bol GH, Kroon PS, et al. Adaptive radiotherapy: the Elekta Unity MR-linac concept. *Clin Transl Radiat Oncol*. 2019;18:54-59.
22. Tustison NJ, Avants BB, Cook PA, et al. N4ITK: improved N3 bias correction. *IEEE Trans Med Imaging*. 2010;29:1310-1320.
23. Willigenburg T, Muinck Keizer DM, Peters M, et al. Evaluation of daily online contour adaptation by radiation therapists for prostate cancer treatment on an MRI-guided linear accelerator. *Clin Transl Radiat Oncol*. 2021;27:50-56.
24. Jaderberg M, Simonyan K, Zisserman A, Kavukcuoglu K. Spatial transformer networks. In: Cortes C, Lawrence N, Lee D, Sugiyama M, Garnett R, eds. *Advances in Neural Information Processing Systems*. Vol 28. Curran Associates, Inc.; 2015.
25. De Vos BD, Berendsen FF, Viergever MA, Sokooti H, Staring M, Išgum I. A deep learning framework for unsupervised affine and deformable image registration. *Med Image Anal*. 2019;52:128-143.
26. Hering A, Hansen L, Mok TCW, et al. Learn2Reg: comprehensive multi-task medical image registration challenge, dataset and evaluation in the era of deep learning. *IEEE Trans Med Imaging*. 2022;42:697-712.
27. Balakrishnan G, Zhao A, Sabuncu MR, Guttag J, Dalca AV. VoxelMorph: a learning framework for deformable medical image registration. *IEEE Trans Med Imaging*. 2019;38:1788-1800.
28. Hendriksen AA. ahendriksenh/msd_pytorch: v0.7.2. 2019. Accessed 2021. https://github.com/ahendriksen/msd_pytorch
29. Minnema J, Eijnatten M, Hendriksen AA, et al. Segmentation of dental cone-beam CT scans affected by metal artifacts using a mixed-scale dense convolutional neural network. *Med Phys*. 2019;46:5027-5035.
30. Avants BB, Epstein CL, Grossman M, Gee JC. Symmetric diffeomorphic image registration with cross-correlation: evaluating automated labeling of elderly and neurodegenerative brain. *Med Image Anal*. 2008;12:26-41.
31. Avants BB, Tustison N, Song G. Advanced normalization tools (ANTS). *Insight J*. 2009;2:1-35.
32. Klein S, Staring M, Murphy K, Viergever MA, Pluim JPW. Elastix: a toolbox for intensity-based medical image registration. *IEEE Trans Med Imaging*. 2009;29:196-205.
33. Pérez-García F, Sparks R, Ourselin S. TorchIO: a Python library for efficient loading, preprocessing, augmentation and patch-based sampling of medical images in deep learning. *Comput Methods Programs Biomed*. 2021;208:106236.
34. Shen Z, Han X, Xu Z, Niethammer M. Networks for joint affine and non-parametric image registration. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2019:4224-4233.
35. Aldoj N, Biavati F, Michallek F, Stober S, Dewey M. Automatic prostate and prostate zones segmentation of magnetic resonance images using DenseNet-like U-net. *Sci Rep*. 2020;10:1-17.
36. Roach D, Holloway LC, Jameson MG, et al. Multi-observer contouring of male pelvic anatomy: highly variable agreement across conventional and emerging structures of interest. *J Med Imaging Radiat Oncol*. 2019;63:264-271.
37. Pathmanathan AU, McNair HA, Schmidt MA, et al. Comparison of prostate delineation on multimodality imaging for MR-guided radiotherapy. *Br J Radiol*. 2019;92:20180948. PMID: 30676772.
38. Thörnqvist S, Petersen JBB, Høyer M, Bentzen LN, Muren LP. Propagation of target and organ at risk contours in radiotherapy of prostate cancer using deformable image registration. *Acta Oncol*. 2010;49:1023-1032.
39. Van Herk M, Bruce A, Kroes AP, Shouman T, Touw A, Lebesque JV. Quantification of organ motion during conformal radiotherapy of the prostate by three dimensional image registration. *Int J Radiat Oncol Biol Phys*. 1995;33:1311-1320.
40. Rueckert D, Aljabar P. Nonrigid registration of medical images: theory, methods, and applications [applications corner]. *IEEE Signal Process Mag*. 2010;27:113-119.
41. Elmahdy MS, Beljaards L, Yousefi S, et al. Joint registration and segmentation via multi-task learning for adaptive radiotherapy of prostate cancer. *IEEE Access*. 2021;9:95551-95568.

SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

How to cite this article: Kolenbrander ID, Maspero M, Hendriksen AA, et al. Deep-learning-based joint rigid and deformable contour propagation for magnetic resonance imaging-guided prostate radiotherapy. *Med Phys*. 2024;1-11. https://doi.org/10.1002/mp.17000