

Feasibility of Image Registration for Ultrasound-Guided Prostate Radiotherapy Based on Similarity Measurement by a Convolutional Neural Network

Technology in Cancer Research & Treatment
Volume 18: 1-11
© The Author(s) 2019
Article reuse guidelines:
sagepub.com/journals-permissions
DOI: 10.1177/1533033818821964
journals.sagepub.com/home/tct


Ning Zhu, PhD^{1,2}, Mohammad Najafi, PhD^{2,3}, Bin Han, PhD^{2,4}, Steven Hancock, MD^{2,4}, and Dimitre Hristov, PhD^{2,4} 

Abstract

Purpose: Registration of 3-dimensional ultrasound images poses a challenge for ultrasound-guided radiation therapy of the prostate since ultrasound image content changes significantly with anatomic motion and ultrasound probe position. The purpose of this work is to investigate the feasibility of using a pretrained deep convolutional neural network for similarity measurement in image registration of 3-dimensional transperineal ultrasound prostate images. **Methods:** We propose convolutional neural network-based registration that maximizes a similarity score between 2 identical in size 3-dimensional regions of interest: one encompassing the prostate within a simulation (reference) 3-dimensional ultrasound image and another that sweeps different spatial locations around the expected prostate position within a pretreatment 3-dimensional ultrasound image. The similarity score is calculated by (1) extracting pairs of corresponding 2-dimensional slices (patches) from the regions of interest, (2) providing these pairs as an input to a pretrained convolutional neural network which assigns a similarity score to each pair, and (3) calculating an overall similarity by summing all pairwise scores. The convolutional neural network method was evaluated against ground truth registrations determined by matching implanted fiducial markers visualized in a pretreatment orthogonal pair of x-ray images. The convolutional neural network method was further compared to manual registration and a standard commonly used intensity-based automatic registration approach based on advanced normalized correlation. **Results:** For 83 image pairs from 5 patients, convolutional neural network registration errors were smaller than 5 mm in 81% of the cases. In comparison, manual registration errors were smaller than 5 mm in 61% of the cases and advanced normalized correlation registration errors were smaller than 5 mm only in 25% of the cases. **Conclusion:** Convolutional neural network evaluation against manual registration and an advanced normalized correlation -based registration demonstrated better accuracy and reliability of the convolutional neural network. This suggests that with training on a large data set of transperineal ultrasound prostate images, the convolutional neural network method has potential for robust ultrasound-to-ultrasound registration.

Keywords

ultrasound image registration, prostate, radiotherapy, image-guided radiation therapy (IGRT), deep convolutional neural network

Abbreviations

ANC, advanced normalized correlation; CNN, convolutional neural network; DRRs, digitally reconstructed radiography; EBRT, external beam radiation therapy; IGRT, image-guided radiation therapy; ReLu, rectified linear unit; US, ultrasound

Received: January 26, 2018; Revised: September 30, 2018; Accepted: November 30, 2018.

¹ Google, Santa Clara County, CA, USA

² Department of Radiation Oncology, Stanford University, Stanford, CA, USA

³ Amazon, Development Engineer II, Seattle, WA, USA

⁴ Department of Radiation Oncology, School of Medicine, Stanford University, Stanford, CA, USA

Corresponding Author:

Ning Zhu, PhD, Google, 1600 Amphitheatre Parkway, Mountain View, Santa Clara County, CA 94043, USA.

Email: zhun@google.com



Creative Commons Non Commercial CC BY-NC: This article is distributed under the terms of the Creative Commons Attribution-NonCommercial 4.0 License (<http://www.creativecommons.org/licenses/by-nc/4.0/>) which permits non-commercial use, reproduction and distribution of the work without further permission provided the original work is attributed as specified on the SAGE and Open Access pages (<https://us.sagepub.com/en-us/nam/open-access-at-sage>).

Introduction

The ability to accurately aim radiation beams at the intended target while avoiding surrounding healthy tissues is critical for the success of prostate external beam radiation therapy (EBRT). Currently, implanted markers are used for accurate prostate localization during EBRT. However, there are several disadvantages with this approach such as morbidity associated with the implantation procedure,¹⁻³ lack of volumetric information for managing anatomic deformations and volume changes,⁴⁻⁶ and potential marker migration before and during radiotherapy that may result in systematic errors.^{1,2,4}

Transperineal ultrasound (US) prostate imaging was recently introduced commercially and deployed clinically^{7,8} as an alternative nonionizing image-guidance modality that could potentially eliminate some of the limitations of transabdominal US guidance.^{9,10} However, US image guidance is challenged by variable operator-dependent image quality and technique-induced nontrivial differences in images of the same anatomy.^{10,11} Intensity-based image registration methods are widely used for medical image registration applications.¹¹⁻¹³ However, due to comparatively low image quality of US images,¹⁴ standard intensity-based similarity metrics for US image registration do not guarantee a satisfactory performance. Furthermore, corresponding 3-dimensional (3-D) US image pairs can appear quite different depending on the transducer position and orientation and thus confound predetermined image features. As a result, intensity-based methods may not be very robust for 3-D US image registration. Even the manual registration of US volumes can be a difficult task.

In this article, we evaluate the feasibility of an alternative approach, a 3-D US image registration framework based on image matching with a pretrained deep convolutional neural network (CNN). Deep CNNs present a powerful methodology that has been used for a variety of medical image analysis tasks,^{15,14} but research on CNNs for medical image registration is still considered to be in early stage¹⁵ with few articles on the subject.¹⁶⁻²¹ For multimodal image registration in particular, an emerging concept is to use CNNs on registered and misregistered image pairs in order to learn and subsequently apply a similarity measure that captures the underlying complex correlation across modalities.^{16,17,20} We consider such CNN-based strategy particularly attractive for the registration of US image pairs acquired at different time instances given that these images generally present nontrivial confounding differences in intensity and content.

Using CNN to measure image similarity ideally requires that a CNN be trained with 3-D US images having ground truth registration results in order to have the CNN design and learn robust US image features most suitable for the application. However, acquiring a large number of US training data sets with validated ground truth registration is logistically challenging. We hypothesize that a pretrained deep CNN²² designed to find correspondence (similarity) of image patches can still be used to measure the similarity of US images as such a network has been trained on a large data set to successfully compare image patches while accounting for a wide variety of changes in image appearance. Thus, we design a registration method based on this pretrained deep CNN and evaluate its performance with 3-D transperineal US images acquired from patients undergoing prostate radiotherapy.

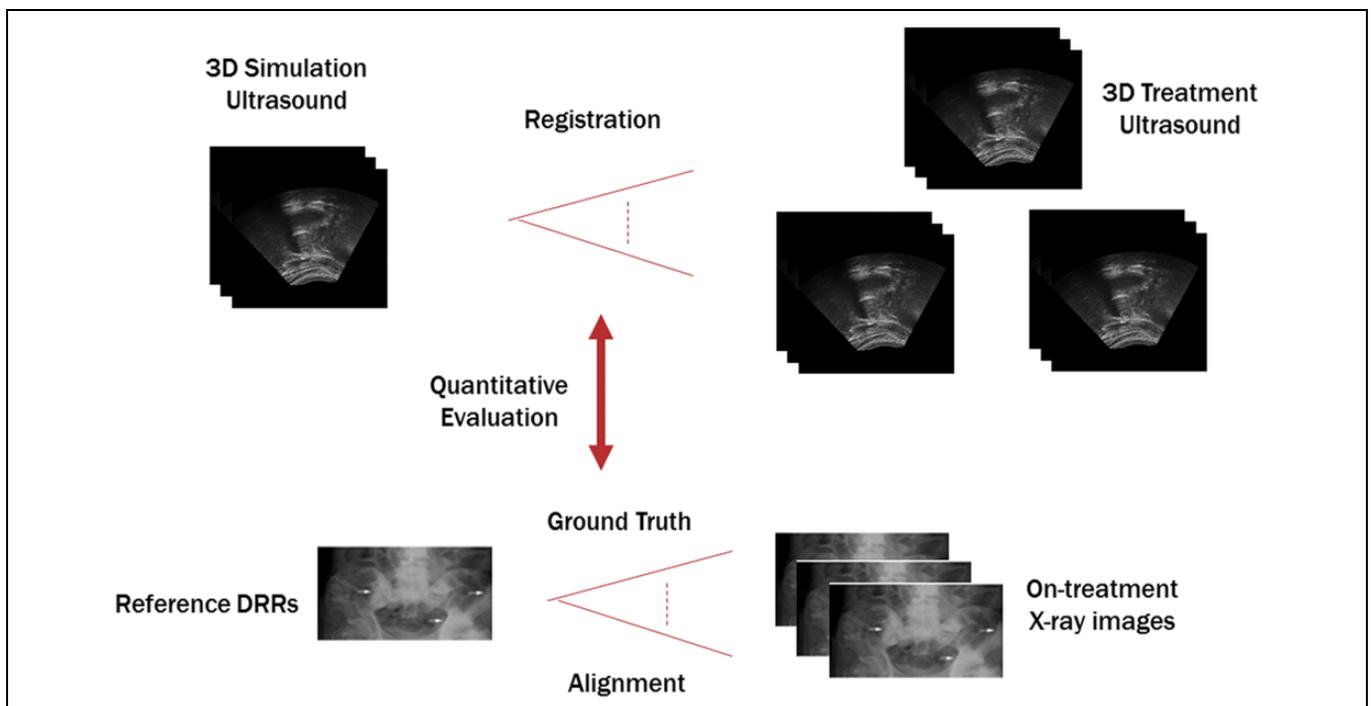


Figure 1. Study design, ground truth, and quantitative evaluation.

Methods and Materials

Treatment Procedure and Data Acquisition

For this study, with institutional review board approval transperineal US imaging of the prostate was performed with the Clarity Autoscan system (Elekta, Stockholm, Sweden) for several prostate patients during simulation and treatment delivery. The Clarity Autoscan system combines infrared tracking and US imaging with the Clarity Autoscan probe to enable prostate localization during radiotherapy simulation and treatment. The Clarity Autoscan US probe is enclosed mechanically swept 3 to 7 MHz transducer that provides 3-D US images through the acquisition of a series of 2-D planes along the elevational direction of the transducer. In particular, for the acquisition of 3-D transperineal US images of the prostate, the probe is placed between the patient’s legs in contact with the perineum. This placement allows prostate imaging through the acoustic window provided by the perineum. The specific data acquisition throughout simulation and treatment is briefly described below.

Prior to computed tomography (CT) simulation scanning, the Clarity US probe is fixed in imaging position between the patient’s legs and left in place throughout the simulation procedure. Infrared reflective markers attached to the probe are tracked by a calibrated camera fixed on the room ceiling. This allows a simulation 3-D US image acquired immediately after the CT simulation scan in the same patient position to be reconstructed and referenced in the coordinate system of the CT device and thus automatically registered to the planning CT. Once completed, the CT contours of several structures (prostate, bladder, and rectum) are transferred from the planning CT to the simulation US image. The prostate contours (with modifications if deemed necessary) are set as an image-guidance volume. Once approved, a treatment plan is imported in the Clarity system to localize the treatment isocenter position within the 3-D simulation US image. (The treatment isocenter

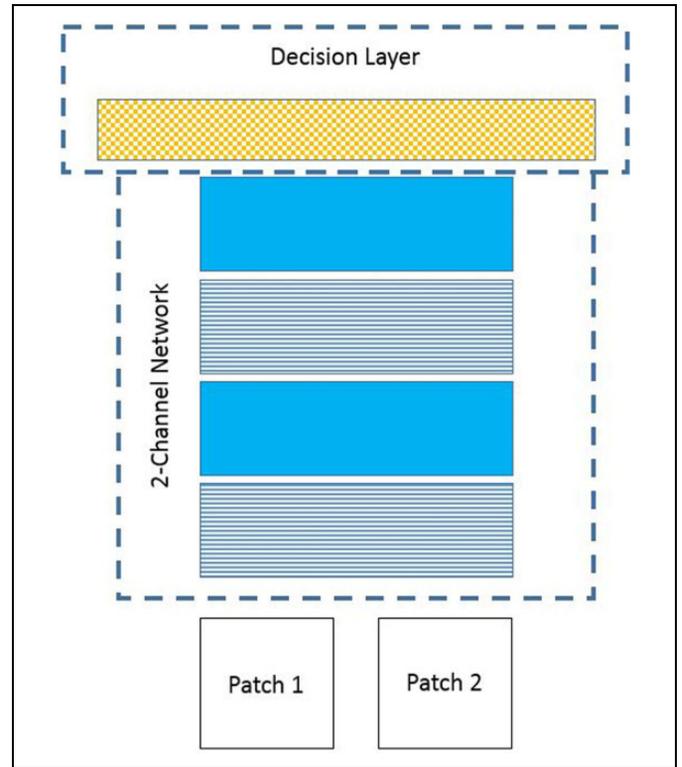


Figure 2. Pretrained deep convolutional neural network used in this study. Pattern code used: Horizontal stripes = Conv + ReLU, solid color = max-pooling, checkered = fully connected later (ReLU exists between fully connected layers as well).²² Conv indicates convolutional neural network; ReLU, rectified linear unit.

is a fixed point in the coordinate system of the medical linear accelerator at the focus of the central axes of all radiation beams deliverable by the accelerator.)

Before treatment, the Clarity US probe is fixed in imaging position between patient’s legs and left in place throughout the

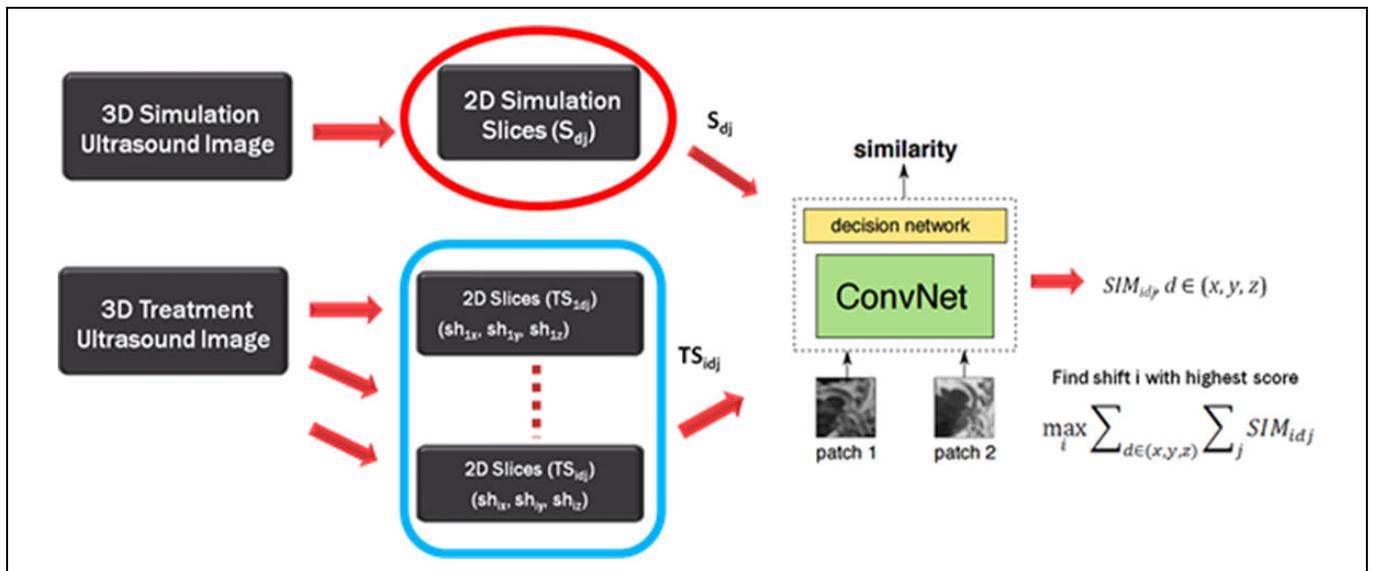


Figure 3. Ultrasound-to-ultrasound registration framework.

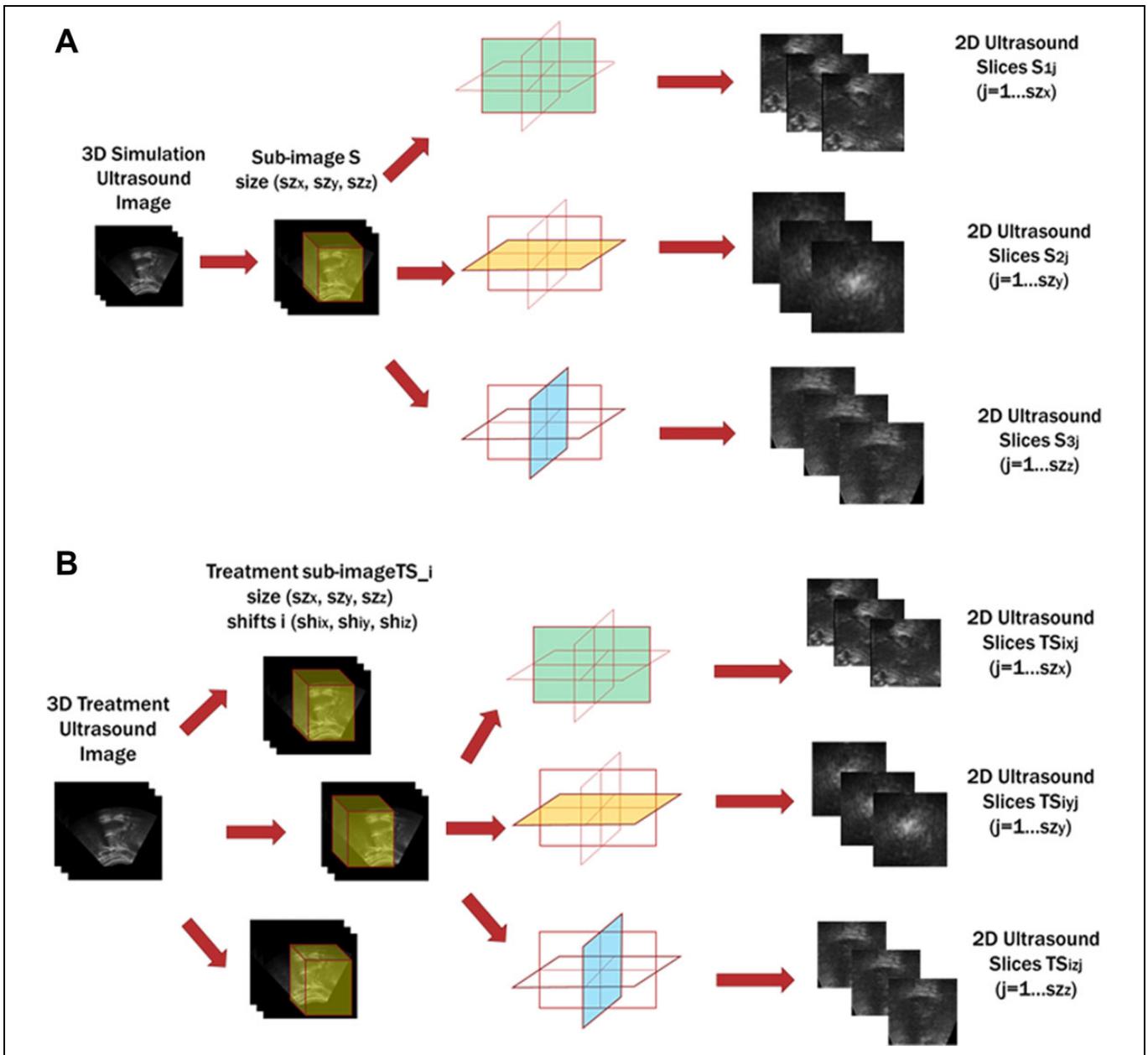


Figure 4. Subimage selection and 2-D image slicing (patch extraction). (A) 2-D slicing of the 3-D simulation US image, (B) 2-D slicing of the 3-D treatment US image. US indicates ultrasound.

treatment procedure including the actual beam delivery. Infrared reflective markers attached to the probe are tracked by a calibrated camera fixed on the room ceiling. This allows a treatment 3-D US image acquired before radiation delivery to be reconstructed and referenced in the coordinate system of the medical linear accelerator. The treatment 3-D US image is registered to the planning 3-D US image manually by overlaying the image-guidance volume (prostate) contours from the simulation US onto the prostate identified on the treatment US. A 3-D shift vector representing a rigid translational transform is then calculated by the Clarity system such that the isocenter-prostate spatial relation reflected in the treatment US image matches the intended isocenter-prostate spatial relation

captured in the planning US image. Hereafter, we refer to the rigid translational transform obtained in this manner as manual registration. The manual registration is recorded but not applied for the treatment.

Commonly, prostate image-guided radiation therapy (IGRT) relies on implanted fiducials to align the prostate target prior to radiation delivery. To this end, as illustrated in Figure 1 (bottom), reference digitally reconstructed radiographies (DRRs) are generated from a CT volume during treatment planning. The DRRs capture the positions of the projected fiducials markers with regard to the treatment isocenter. Thus concurrently with the treatment US acquisition and registration, a pair of 2-D x-ray images are acquired with an On-

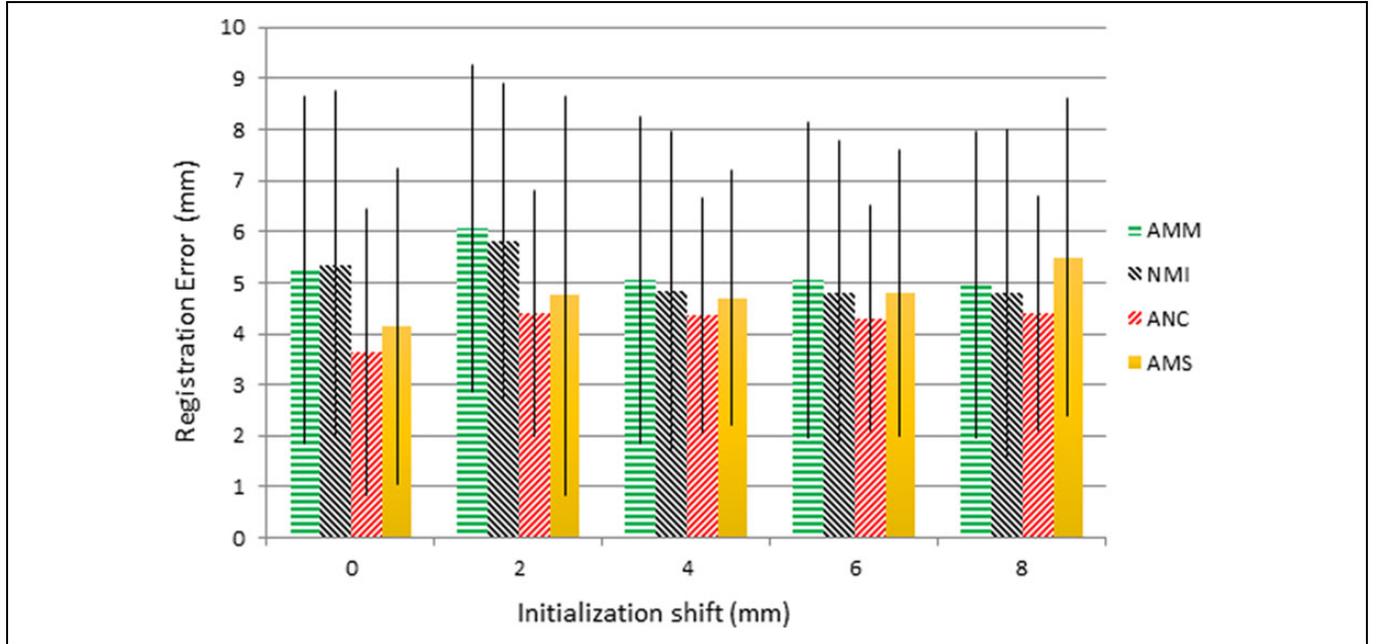


Figure 5. Mean registration errors for different similarity metrics on developmental images.

Board Imager on a Varian 23EX Linac (Varian Medical Systems, Palo Alto, California). Such a pair of 2-D x-ray images allows localization of the fiducials in the coordinate system of the radiation delivery system. Then, a rigid body 3 degree-of-freedom transform (a 3-component vector) is calculated by aligning 4 prostate-implanted fiducial markers in corresponding pairs of reference DRRs and the 2-D x-ray images. This 3-D shift vector represents the rigid translational transformation that needs to be applied to match the isocenter-prostate spatial relation captured by the pair of x-ray images to the intended isocenter-prostate spatial relation captured in the simulation CT. Ideally, both the US and the x-ray image guidance should result in the same prostate shifts to align the target in the coordinate space of the treatment device. Discrepancies are interpreted as errors in the US-US registration in comparison the x-ray fiducial-based registration that is widely used clinically.

In the present study, the x-ray-based translational transforms (shifts) calculated for patients undergoing prostate IGRT serve as a ground truth for evaluating the accuracy of the proposed US-to-US registration method. Simulation 3-D US images acquired during initial planning CT and 3-D US treatment images (US images acquired right before treatment) serve as inputs and a translational transform (vector shift) is the output as shown in Figure 1 (top). The evaluation is conducted by calculating the norm of the difference between the 2 registration vectors (ground truth and the results obtained with the proposed method).

Deep CNN

In the proposed method, a CNN (or ConvNet) is used for matching of 2-D image slices. Convolutional neural network is a type of feed-forward artificial neural network in machine learning that is proven to be successful for image and video

analysis. The input for the network is an image pair of 2-D slices and the output is a similarity score. Due to the lack of training data for the deep CNN, the proposed method uses a pretrained CNN (Figure 2) described by Zagoruyko *et al.*²²

The pretrained CNN (Figure 2) designed to find correspondence (similarity) of image patches consists of convolutional layers, rectified linear unit (ReLU) layers, max-pooling layers, and a fully connected layer (for overview of CNNs architectures, refer to²³ and references therein). Specifically, a list of all layers from bottom up includes convolutional layer 1 ($C(96, 7, 3)$); ReLU layer 1, max-pooling layer ($P(2, 2)$); convolutional layer 2 ($C(192, 5, 1)$); ReLU layer 2, max-pooling layer 2 ($P(2, 2)$); convolutional layer 3 ($C(256, 3, 1)$); ReLU layer 3, fully connected layer 1 ($F(256)$), and ReLU layer 4, fully connected layer 2 ($F(1)$). Following the notation in a previous study,²² $C(n, k, s)$ is a convolutional layer with n filters of spatial size $k \times k$ applied with stride s , ($P(k, s)$) is a max-pooling layer of size $k \times k$ applied with stride s , and $F(n)$ denotes a fully connected linear layer with n output units.

The output of the network is the output of the fully connected layer ($F(1)$), which is a score number representing the similarity of the 2 input 2-D image slices. The CNN we used is pretrained with the Liberty benchmark data set containing more than 450°000 image patches (64×64 pixels).²² The training process optimizes an objective function with hinge-based loss term and squared l_2 -norm regularization using supervised training with neural network. More training details can be found in the original article.²²

Registration Framework

Figure 3 illustrates the CNN (ConvNet) framework for registering 3-D treatment US images (acquired right before

treatment) to 3-D simulation US images (acquired before planning). Two-D slices (patches) are extracted from the 3-D simulation and treatment US images along the axes d_j ($j = 1, 2, 3$) in world (room) coordinate system.

For each 3-D shift i , a translated treatment 3-D US image is generated. Since a shift i is not necessarily an integer value of the intervoxel spacing, a trilinear interpolation is used to calculate the voxel values of the translated image. A composite similarity score is then calculated by summing up the similarity scores of spatially corresponding patches. The similarity score for each patch pair is calculated with the pretrained ConvNet and a composite similarity score is calculated across all patches. The translational shift that generates the maximum composite similarity score is considered to be the translational transform that best matches treatment and simulation US images. The calculation of similarity score is defined in Equation 1. Figure 4A further details the process of extracting the 2-D slices (patches) from the US simulation image (as indicated by the ellipse in Figure 3), and Figure 4B details the process of extracting the 2-D slices (patches) from shifted treatment images (as indicated by the rectangle in Figure 3).

As shown in Figure 4A, the simulation US image is cropped into a subimage, \mathcal{S} , according to a region of interest $[R_x : R_x + sz_x, R_y : R_y + sz_y, R_z : R_z + sz_z]$ encompassing the prostate. By cropping images into smaller subregions of interest, the tendency of matching images along sector boundary is eliminated. The computational efficiency of the registration is also improved. Here $[R_d : R_d + sz_d]$ is the range along axis d in world (room) coordinate system. The subimage \mathcal{S} with size (sz_x, sz_y, sz_z) is then cut into 3 groups of 2-D slices $(\mathcal{S}_{d_j}, d = x, y, z; j = 1, \dots, sz_d)$ in planes perpendicular to image axes d_j .

The treatment US image is cropped into treatment subimages TS_i corresponding to various shifts for the region of interest. The region of interest encompasses the whole prostate area. The process of cutting and shifting the treatment image is presented in Figure 4B. Assuming the i -th shift vector is $(sh_{ix}, sh_{iy}, sh_{iz})$, the subimage TS_i associated with the i -th shift corresponds to a region of interest $[R_x + sh_{ix} : R_x + sz_x + sh_{ix}, R_y + sh_{iy} : R_y + sz_y + sh_{iy}, R_z + sh_{iz} : R_z + sz_z + sh_{iz}]$. The shift vectors, indexed by i $(sh_{ix}, sh_{iy}, sh_{iz})$, can cover all possible registration shifts. For example, the registration shifts along each axis can range from -15 mm to 15 mm, then the shifts sh_{id} for $d = x, y, z$ are set to integer shifts (in millimeter) within the range. To accelerate the calculation procedure, a multi-scale method is used. The spacing for the shifts are set to sp_k in the k -th stage, $k = 1, 2$. The spacing of the first stage is set to be larger than the second stage, so as to accelerate the registration process. The search range for the second stage is set around the shifts with maximum similarity response in the first stage. For each shifted subimage, the patch extraction procedure is similar to that of the simulation subimage. After obtaining the 3 groups of 2-D slices for both the simulation subimage $(\mathcal{S}_{d_j}, d = x, y, z; j = 1, \dots, sz_d)$ and shifted treatment subimage $(TS_{id_j}, d = x, y, z; j = 1, \dots, sz_d)$, the corresponding

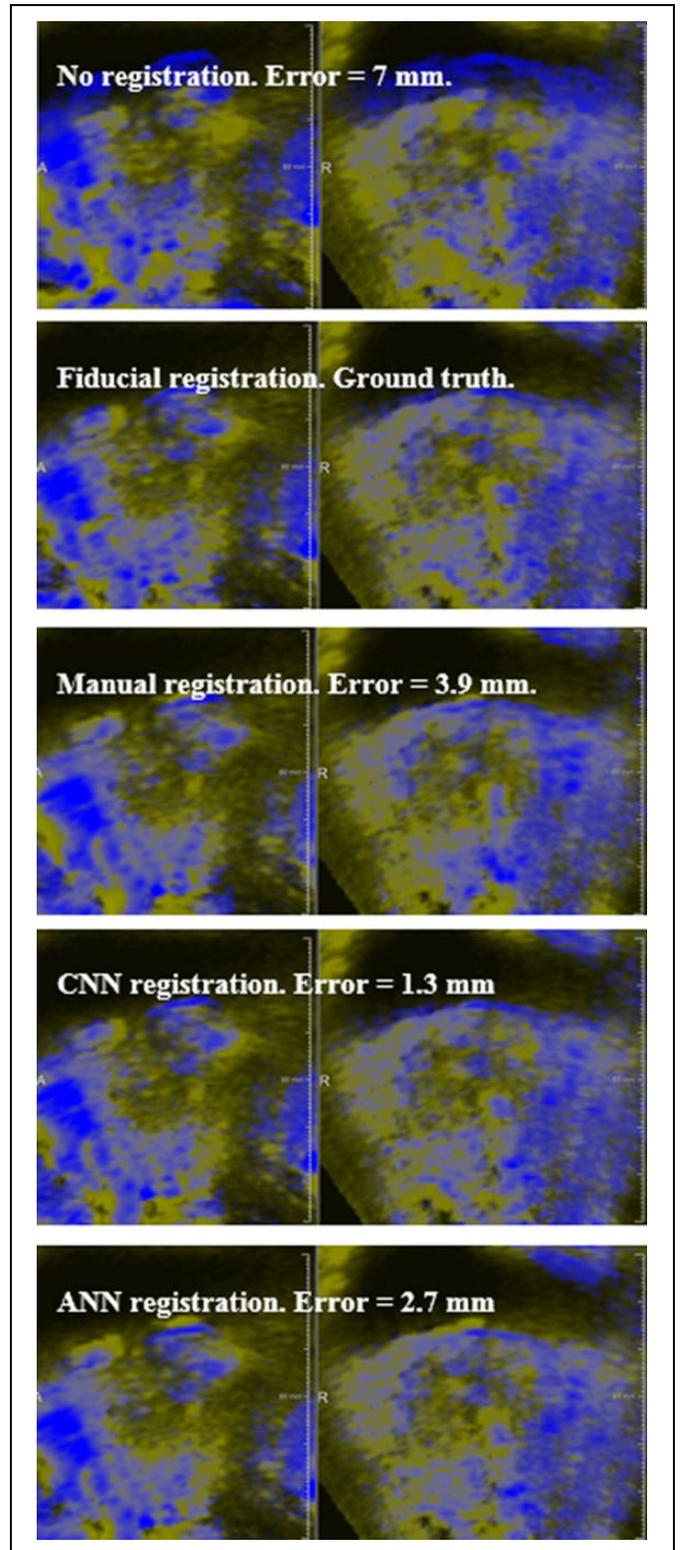


Figure 6. Sagittal (left) and coronal (right) planes of a simulation ultrasound image (in yellow) and a treatment ultrasound image (in blue) overlaid after registration with various methods. The x-ray-based fiducial registration serves as ground truth. The reported registration error is the norm of the difference between 2 vectors: the vector for the ground truth shift and the vector for the respective evaluated registration.

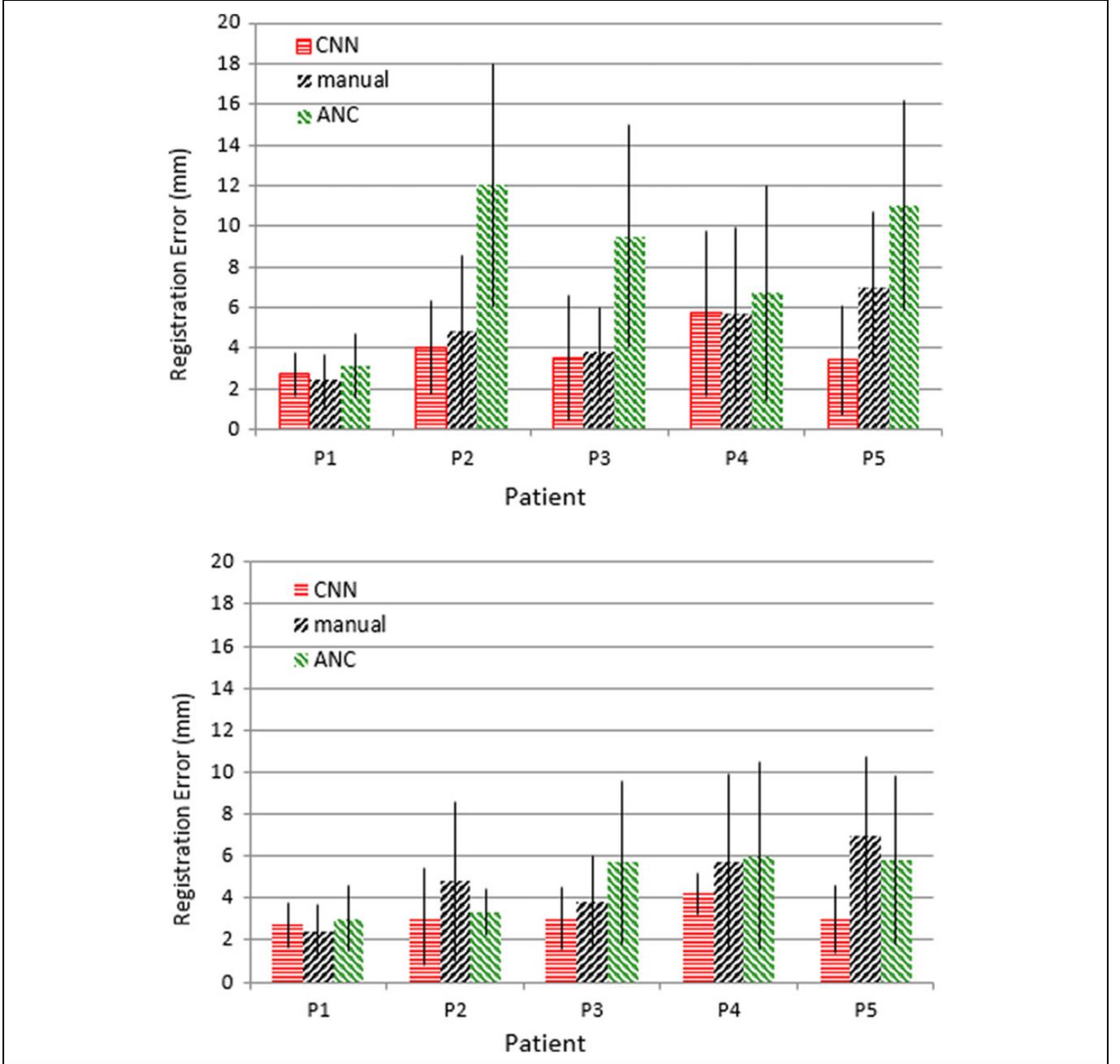


Figure 7. Performance comparison of different US–US registration methods: proposed (CNN), manual registration, and ANC (Elastix). Top: Mean errors without registration initialization. Bottom: Mean errors with registration initialization. In this case, CNN (proposed) and ANC registrations are performed starting with a randomly selected 4 mm initial shift from the ground truth registration. US indicates ultrasound; CNN, convolutional neural network; ANC, advanced normalized correlation.

2-D slice pairs (S_{dj} and TS_{idj}) are used as input pairs to the pretrained CNN.

The output of the network for a 2-D slice pair is a similarity SIM_{idj} . Then for each shifted treatment subimage $TS_i (i = 1, \dots, N)$, the sum of the similarities of all 3 groups of 2-D slices is:

$$SIM_i = \sum_{d=x,y,z} \sum_{j=1}^{sz_d} SIM_{idj} \quad (1)$$

After obtaining all SIM_i for the shifted treatment subimages $TS_i (i = 1, \dots, N)$, the result of the registration is determined by choosing the shift i with the highest score.

Evaluation

We compared the proposed method to results from manual registrations and some of the popular standard intensity-

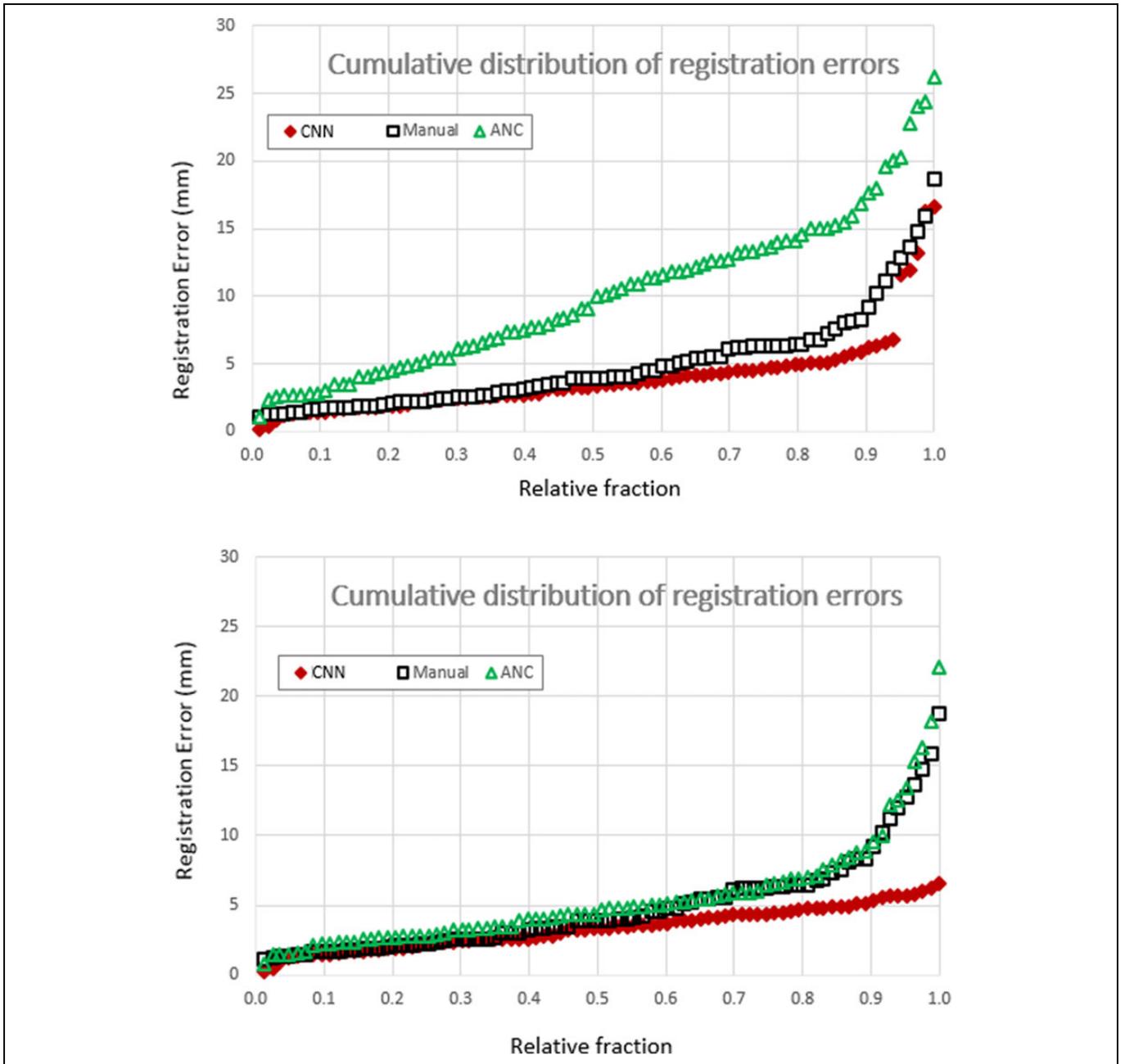


Figure 8. Cumulative distributions of registration errors for the proposed (CNN), manual, and ANC registration methods. Top: Without initialization. Bottom: With initialization. CNN indicates convolutional neural network; ANC, advanced normalized correlation.

based similarity metrics. We used Elastix^{24,25} which is a widely used image registration tool with multiple choices of similarity metrics as an implementation of intensity-based registration.

In our experiments, 121 3-D US images from 5 patients (P1-P5) were used for development and validation. The 3-D US images were available from the Clarity system at up-sampled uniform voxel size of $0.58 \text{ mm} \times 0.58 \text{ mm} \times 0.58 \text{ mm}$. (The inherent resolution of US images acquired with the Clarity abdominal transducer is about 0.5 mm in axial [along beam propagation], 2 mm in lateral [within imaging plane], and 4 mm

in elevational direction). The data set for development consisted of 38 images from the first 3 patients (P1, P2, and P3). It was used to (1) find the similarity metric with the best performance using the Elastix implementation and (2) to identify the sum of CNN generated similarity scores $SIM_i = \sum_{d=x,y,z} \sum_{j=1}^{SZ_d} SIM_{i,dj}$ between 2-D patches (see Equation 1) as the combination leading to best performance of the proposed registration framework. The developmental data set was chosen as the first half (in chronological order) of each patient's images. To determine the best performance similarity metrics for 3-D US image registration, a series of experiments with

different similarity metrics and different shift initialization values were conducted. Four popular similarity metrics advanced (Mattes mutual information), normalized mutual information, advanced normalized correlation (ANC), and advanced mean squares) were tested. Since the ground truth registration for each developmental treatment image was known, the initializations were set at 0, 2, 4, 6, and 8 mm away from the ground truth. In this manner, we examined the changes of performance for a particular similarity metric with changes of initializations. Elastix parameters other than the similarity metrics were set as follows: “NumberOfHistogramBins” = 32, “MaximumNumberOfIterations” = 250, and “NumberOfSpatialSamples” = 2048. Another important Elastix parameter is “NumberOfResolutions.” Values from 1 to 5 were tested and the final value was set to 4 as it provided the best Elastix results in the developmental experiments. The spatial transform in Elastix was restricted to 3-D rigid body translation.

Results

Figure 5 presents mean registration errors and respective standard deviations for different similarity metrics used with Elastix on 38 developmental images and different initialization values for the shifts. The ANC appeared to slightly outperform the other metrics in terms of mean values and spread (as determined by the standard deviation). Thus, the ANC metric was subsequently used with Elastix.

Figure 6 illustrates registrations performed by the 3 evaluated methods (manual, CNN, ANC) along with the ground truth registration and the starting (no registration) point for the registrations. Figure 6 exemplifies the challenge in interpreting the similarity between US images by clearly demonstrating that even in a case where visually the manual, CNN, and ANC methods performed reasonably well, the registration error varied substantially between them.

The CNN method was then compared to manual registrations (as performed by physicists at the time of treatment) and Elastix with ANC. In the performance comparison, 83 images (second half of the P1-P3 data sets and P4-P5 complete data sets) were used for the evaluation.

Figure 7 illustrates mean errors and respective standard deviations for the 3 evaluated registration methods. Figure 7 (top) presents results without initialization and Figure 7 (bottom) presents results with initialization. The initialization shift was chosen as a random vector of size 4 mm away from the ground truth. Without initialization, the ANC registration performs poorly in comparison to the other methods both in terms of mean errors and standard deviations (Figure 7, top). Without initialization, the CNN performance was comparable to or better than manual registration (Figure 7, top).

With initialization around the ground truth, ANC performance improved (Figure 7, bottom) but remained inferior to that of CNN both in terms of mean errors and standard deviations. With initialization, the performance of the proposed

CNN method remained comparable or better to manual registrations both in terms of mean errors and standard deviations.

Figure 8 presents the cumulative distributions of the registration methods across all 5 patients on the 83 validation image pairs. It demonstrates that with initialization in 88% of the cases CNN registration errors were smaller than 5 mm. Without initialization, in 81% of the cases CNN registration errors were smaller than 5 mm. The corresponding values for the ANC method were 62% and 25% accordingly, whereas for the manual registration these were within 5 mm in 61% of the cases. These results clearly demonstrate the improvement in overall registration accuracy that can be achieved with a pretrained CNN in comparison to standard manual or automatic intensity-based registration techniques.

Discussion and Conclusion

In this article, we designed and evaluated a 3-D US image registration framework based on a pretrained deep CNN. Comparative evaluation of the method against manual registration and automatic intensity-based registration with an ANC similarity metric demonstrated significantly improved accuracy and reliability with the pretrained CNN approach. One limitation of the study is that the registration transformation had to be limited to translations only since the available “ground truth” registrations were 3-D translations obtained by x-ray-based marker matching performed by treating therapists. Standard intensity-based registrations may perform better if deformations are considered and this scenario should be the subject of further investigations.

Our results on the accuracy of the pretrained deep CNN approach to US-US registration need to be interpreted in the context of several uncertainties related to the establishment of the “ground truth” x-ray-based image registration. Prostate deformations, for instance, may be present between simulation and treatment due to differences in rectal and bladder filling as well as probe pressure. The magnitude of these deformations is patient and session dependent. We evaluated the prostate distortions by measuring the relative changes in interfiducial distances from simulation to treatment. On average, the relative change was smaller than 2% or 0.5 mm for mean interfiducial distance of 25 mm.

Uncertainty in marker localization arising from user bias in x-ray image interpretation is another source of error in the determination of the ground truth. We evaluated this by comparing the x-ray-based shifts that we calculated to the shifts approved and applied by the therapists during the actual treatments. The standard deviation of the difference vector was (0.6, 0.6, 0.5) mm, resulting in approximately 2 mm overall uncertainty at the 95% confidence level. This number provides an estimate of the ground truth error in our study.

Our results indicate clearly the potential of using deep CNNs for 3-D US image registration, but the overall accuracy of the current approach based on a specific, pretrained CNN is not sufficient to meet the requirements of prostate IGRT even after considering uncertainty in ground truth registrations. This

is not surprising as the CNN was pretrained with nonmedical image data. Hence, it is expected that training the CNN with actual US data can notably enhance the CNN performance and future work will involve network training on a large data set of US images. Furthermore, for practical implementation additional performance optimizations will be necessary. On our hardware, it takes about 5 milliseconds to compute the CNN similarity between a pair of 64×64 2-D patches. Thus, about 1 second is necessary to calculate the similarity between a pair of 3-D images, as this involves $64 * 3 = 192$ evaluations between 2-D patches. In comparison, normalized mutual information computation took about 5 milliseconds. A straightforward optimization, for instance, would be to reduce the number of patches used for composite similarity measurement to only few that are rich in relevant anatomical features.

We expect that performance optimizations and training application-specific US images will allow CNN-based registration to address robustly the challenge of US-to-US prostate registration and eliminate a major obstacle for US IGRT.

Authors' Note

The data collected in this study were acquired under Stanford IRB-approved protocol #27372 "Feasibility of using trans-perineal Clarity Autoscan ultrasound imaging for prostate motion management, tissue characterization, and treatment monitoring." Patients accrued under this protocol provided written consent for participation in the study and publication of the findings. Our study was approved by the Stanford University Research Compliance Office, IRB number 27372. All patients provided written informed consent prior to enrollment in the study. TCRT-18-0047.R3.

Declaration of Conflicting Interests

The author(s) declared the following potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: The corresponding author received research funding from Philips and Elekta.

ORCID iD

Dimitre Hristov, PhD  <https://orcid.org/0000-0002-2462-7727>

References

- Shirato H, Harada T, Harabayashi T, et al. Feasibility of insertion/implantation of 2.0-mm-diameter gold internal fiducial markers for precise setup and real-time tumor tracking in radiotherapy. *Int J Radiat Oncol Biol Phys.* 2003;56(1):240-247.
- Langenhuijsen JF, van Lin EN, Kiemeny LA, et al. Ultrasound-guided transrectal implantation of gold markers for prostate localization during external beam radiotherapy: complication rate and risk factors. *Int J Radiat Oncol Biol Phys.* 2007;69(3):671-676.
- Henry AM, Wilkinson C, Wylie JP, Logue JP, Price P, Khoo VS. Trans-perineal implantation of radio-opaque treatment verification markers into the prostate: an assessment of procedure related morbidity, patient acceptability and accuracy. *Radiother Oncol.* 2004;73(1):57-59.
- Nichol AM, Brock KK, Lockwood GA, et al. A magnetic resonance imaging study of prostate deformation relative to implanted gold fiducial markers. *Int J Radiat Oncol Biol Phys.* 2007;67(1):48-56.
- Kupelian PA, Willoughby TR, Meeks SL, et al. Intraprostatic fiducials for localization of the prostate gland: monitoring intermarker distances during radiation therapy to test for marker stability. *Int J Radiat Oncol Biol Phys.* 2005;62(5):1291-1296.
- Ghilezan MJ, Jaffray DA, Siewerdsen JH, et al. Prostate gland motion assessed with cine-magnetic resonance imaging (cine-MRI). *Int J Radiat Oncol Biol Phys.* 2005;62(2):406-417.
- Baker M, Behrens CF. Determining intrafractional prostate motion using four dimensional ultrasound system. *BMC Cancer.* 2016;16:484.
- Baker M, Behrens CF. Prostate displacement during transabdominal ultrasound image-guided radiotherapy assessed by real-time four-dimensional transperineal monitoring. *Acta Oncol.* 2015;54(9):1508-1514.
- McGahan JP, Ryu J, Fogata M. Ultrasound probe pressure as a source of error in prostate localization for external beam radiotherapy. *Int J Radiat Oncol Biol Phys.* 2004;60(3):788-793.
- Artignan X, Smitsmans MH, Lebesque JV, Jaffray DA, van Her M, Bartelink H. Online ultrasound image guidance for radiotherapy of prostate cancer: impact of image acquisition on prostate displacement. *Int J Radiat Oncol Biol Phys.* 2004;59(2):595-601.
- Zitova B, Flusser J. Image registration methods: a survey. *Image Vis Comput.* 2003;21:977-1000.
- Hill DL, Batchelor PG, Holden M, Hawkes DJ. Medical image registration. *Phys Med Biol.* 2001;46(3):R1-R45.
- Maintz JB, Viergever MA. A survey of medical image registration. *Med Image Anal.* 1998;2(1):1-36.
- Shen D, Wu G, Suk HI. Deep learning in medical image analysis. *Annu Rev Biomed Eng.* 2017;19:221-248.
- Litjens G, Kooi T, Bejnordi BE, et al. A survey on deep learning in medical image analysis. *Med Image Anal.* 2017;42:60-88.
- Cheng X, Zhang L, Zheng Y. Deep similarity learning for multimodal medical images. *Computer methods in biomechanics and biomedical engineering. Imaging Vis.* 2018;6:248-252.
- Wang SY, Kim M, Wu GR, Shen DG. Scalable high performance image registration framework by unsupervised deep feature representations learning. *Els Mic Soc Book Ser.* 2017;11:245-269.
- de Vos B, Berendsen F, Viergever M, Staring M, Isgum I. End-to-end unsupervised deformable image registration with a convolutional neural network. *Lect Notes Comput Sci.* 2017;10553:204-212.
- Yang X, Kwitt R, Niethammer M. Fast Predictive Image Registration. *Lect Notes Comput Sci.* 2016;10008:48-57.

20. Simonovsky M, Gutierrez-Becker B, Mateus D, Navab N, Komodakis N. A deep metric for multimodal registration. *Lect Notes Comput Sci.* 2016;9902:10-18.
21. Miao S, Wang ZJ, Liao R. A CNN Regression approach for real-time 2D/3D registration. *IEEE Trans Med Imaging.* 2016;35(5):1352-1363.
22. Zagoruyko S, Komodakis N. Learning to compare image patches via convolutional neural networks. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*; June 7–12, 2015; New Jersey, USA: IEEE Publishing; 4353-4361.
23. LeCun Y, Bengio Y, Hinton G. Deep learning. *Nature.* 2015; 521(7553):436-444.
24. Shamonin DP, Bron EE, Lelieveldt BP, Smits M, Klein S, Staring MI. Alzheimer's disease neuroimaging. Fast parallel image registration on CPU and GPU for diagnostic classification of Alzheimer's disease. *Front Neuroinform.* 2013;7:50.
25. Klein S, Staring M, Murphy K, Viergever MA, Pluim JP. Elastix: a toolbox for intensity-based medical image registration. *IEEE Trans Med Imaging.* 2010;29(1):196-205.