Physics in Medicine & Biology

IPEM Institute of Physics and Engineering in Medicine

CrossMark

PAPER

Fully automatic catheter segmentation in MRI with 3D convolutional neural networks: application to MRI-guided gynecologic brachytherapy

Paolo Zaffino[1,8], Guillaume Pernelle[2], Andre Mastmeyer[3], Alireza Mehrtash[4], Hongtao Zhang[5], Ron Kikinis[4,6], Tina Kapur[4,7] and Maria Francesca Spadea[1,7]

[1] Department of Experimental and Clinical Medicine, Magna Graecia University, 88100, Catanzaro, Italy
[2] Bioengineering Department, Imperial College London, SW7 2AZ, London, United Kingdom (at the time of study)
[3] Institute of Medical Informatics, University of Luebeck, 23562, Lübeck, Germany
[4] Department of Radiology, Brigham and Women's Hospital, Harvard Medical School, Boston, MA 02115, United States of America
[5] Oncology Department, Hebei General Hospital, 050051 Shijiazhuang, People's Republic of China
[6] Department of Computer Science at the University of Bremen, Fraunhofer MEVIS, 28359, Bremen, Germany
[7] Both authors contributed equally to this work.
[8] Author to whom any correspondence should be addressed.

E-mail: p.zaffino@unicz.it

Abstract

External-beam radiotherapy followed by high dose rate (HDR) brachytherapy is the standard-of-care for treating gynecologic cancers. The enhanced soft-tissue contrast provided by magnetic resonance imaging (MRI) makes it a valuable imaging modality for diagnosing and treating these cancers. However, in contrast to computed tomography (CT) imaging, the appearance of the brachytherapy catheters, through which radiation sources are inserted to reach the cancerous tissue later on, is often variable across images. This paper reports, for the first time, a new deep-learning-based method for fully automatic segmentation of multiple closely spaced brachytherapy catheters in intraoperative MRI.

Represented in the data are 50 gynecologic cancer patients treated by MRI-guided HDR brachytherapy. For each patient, a single intraoperative MRI was used. 826 catheters in the images were manually segmented by an expert radiation physicist who is also a trained radiation oncologist. The number of catheters in a patient ranged between 10 and 35. A deep 3D convolutional neural network (CNN) model was developed and trained. In order to make the learning process more robust, the network was trained 5 times, each time using a different combination of shown patients. Finally, each test case was processed by the five networks and the final segmentation was generated by voting on the obtained five candidate segmentations. 4-fold validation was executed and all the patients were segmented.

An average distance error of 2.0 ± 3.4 mm was achieved. False positive and false negative catheters were 6.7% and 1.5% respectively. Average Dice score was equal to 0.60 ± 0.17. The algorithm is available for use in the open source software platform 3D Slicer allowing for wide scale testing and research discussion. In conclusion, to the best of our knowledge, fully automatic segmentation of multiple closely spaced catheters from intraoperative MR images was achieved for the first time in gynecological brachytherapy.

Introduction

Gynecologic (GYN) malignancies, which include cervical, endometrial, ovarian, vaginal and vulvar cancers, cause significant mortality in women worldwide. According to an estimation, in the U.S., approximately 110 070 women developed gynecologic cancer in 2018, and over 30% died from the disease (American Cancer Society 2018). The standard-of-care treatment for many primary and recurrent gynecologic cancers consists of external-beam radiation followed by high dose rate (HDR) brachytherapy (Viswanathan *et al* 2012). In contrast to external-beam radiation treatment, in which a linear accelerator aims radiation beams at the pelvis from outside the body, in HDR brachytherapy, radioactive sources, that deliver very high doses of radiation (typically in the range of 80–90 Gray for cervical cancer), are placed directly inside the cancerous tissue using intracavitary or interstitial catheters. The use of imaging to assist gynecologic brachytherapy treatment planning and dose delivery has evolved from two-dimensional (2D) plain x-ray radiographs or ultrasound (US) images to three-dimensional (3D) volumes, including those generated by computed tomography (CT) and magnetic resonance imaging (MRI) (Viswanathan *et al* 2011). Unlike 2D imaging, with 3D imaging, the applicator position and tumor may be visualized and contoured, permitting accurate placement of applicator and tailoring the radiation dose (Fellner *et al* 2001, Haie-Meder *et al* 2005, Pötter *et al* 2006).

Increasing role of MRI in gynecologic cancer treatment

MRI is used routinely in the diagnosis of cervical cancer due to the high contrast of soft tissue which allows differentiation between tumors of the cervix or endometrium from the normal uterus or cervix. Enhanced soft-tissue contrast in MR imaging facilitates the contouring of tumor and organs at risk. This, in turn, facilitates successful image-guided radiation treatment which relies on these contours to maximize radiation dose to the tumor while minimizing dose to organs at risk. In a survey by the American Brachytherapy Society, the utilization of MRI for gynecologic brachytherapy treatment planning increased from 2% to 34% between 2007 and 2014 (Grover *et al* 2016). This development has been driven primarily by the superior soft tissue contrast provided by MRI, and also due to the increased access, availability and familiarity of physicians with MRI. The role of MRI in guidance of brachytherapy applicator placement has also steadily advanced from being feasible in both low-field 0.5 T scanners (Dimopoulos *et al* 2006, Viswanathan *et al* 2006) and high-field 3 T scanners (Kapur *et al* 2012). MRI-guidance demonstrated improvements over CT-guidance in local control and overall survival rates in women with locally advanced cervical cancer (Kamran *et al* 2017). This technological advance has been implemented recently via dedicated MRI-guided gynecologic brachytherapy suites in academic hospitals all over the world (Viswanathan *et al* 2013, Anderson *et al* 2018).

In GYN-HDR interstitial brachytherapy, several catheters (in the investigated dataset up to 35) with tungsten alloy needle inlays (ProGuide Needles, Nucletron Co., Veenendaal, the Netherlands), are inserted percutaneously through a standardized template such as the Syed-Neblett-template (figure 1).

The template is surgically sutured to the patient's perineum and its insertion holes are identified by coordinates.

After insertion, needles are removed and the hollow outer catheters are left in place to serve as channels for delivering radiation sources to the target tissue. Then, MR images are acquired for treatment planning. While the cancer and surrounding soft tissues appear rather distinctively in the MRI, the dark and diffuse appearance of the catheters poses a challenge for both human operators and automatic detection algorithms. Furthermore, the cylindrical plastic obturator, which is commonly used to stabilize the template, creates another signal void that renders some of the catheters on its side partially indistinguishable. Blood vessels as well can be confused with catheters. An example case is provided in figure 2. In the current clinical practice, a CT scan is also acquired since it is easier to manually identify catheters on the CT images. Then, CT and MR images are registered to localize catheters on MRI. This procedure is prone to mis-registration errors, extra scanning time and additional patient's radiation exposure.

State-of-the-art in segmentation methods for catheters, needles, or tubular structures in images

Several approaches have been suggested in the medical image computation literature for segmenting tubular objects. Segmentation of tortuous and tree-like structures, such as blood vessels (Wink *et al* 2004, Liskowski and Krawiec 2016, Mastmeyer *et al* 2016), white matter tracts (O'Donnell and Westin 2007, Hao *et al* 2014) or nerves (Sultana *et al* 2017) are covered by many of the reported methods. Other methods address straight or bent needles or wires (Pernelle *et al* 2013, Hrinivich *et al* 2017, Mastmeyer *et al* 2017). Based on the clinical application, the proposed techniques have been applied to different image modalities including US (Aboofazeli *et al* 2009, Beigi *et al* 2015, Hrinivich *et al* 2017), CT (Görres *et al* 2014, Nguyen *et al* 2015), and MRI (Pernelle *et al* 2013, Mastmeyer *et al* 2017) for the purpose of localization after insertion or real-time guidance during insertion. Many attempts have been made to incorporate hand-crafted and kernel-based methods for detection of lines in images. The reported methods are often based on enhanced ideas of the 3D generalized Hough transform (Beigi
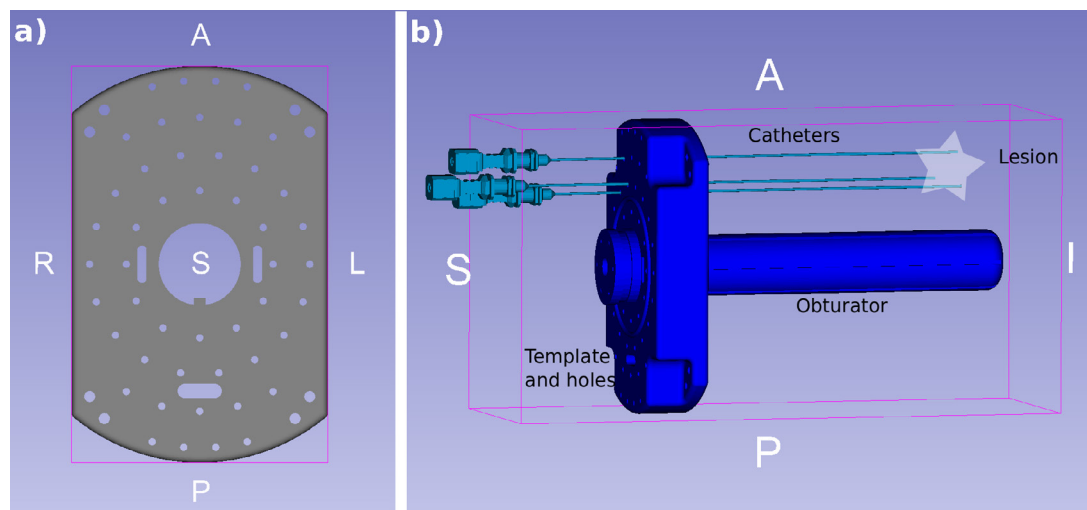
Figure 1. Syed-Neblett-template sutured to the patient's perineum. The holes in the template base (panel (a)) serve as entry points for the brachytherapy catheters and later radiation sources targeting the lesion around the obturator shaft (panel (b)). A similar figure appears in Mastmeyer *et al* (2017).
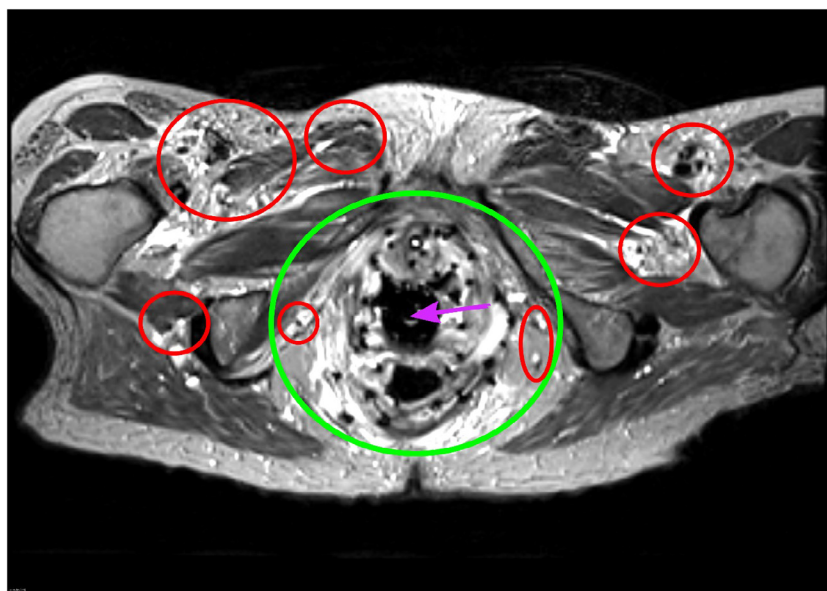

Figure 2. Example case of gynecologic brachytherapy intraoperative MRI. The green circle delimits the area where 35 catheters were placed. The red circles highlight areas containing anatomical structures (e.g. blood vessels) that appear very dark, similar to catheters and can be mistaken for catheters both by operator and by automatic algorithms. Purple arrow points to the center of the obturator shaft area.

et al 2015, Hrinivich *et al* 2017), on model and ray casting based searches (Pernelle *et al* 2013, Mastmeyer *et al* 2017), orthogonal 2D projections (Aboofazeli *et al* 2009), generalized radon transforms (Novotny *et al* 2007, Aboofazeli *et al* 2009) or RANdom Sample Consensus (RANSAC) (Uherčík *et al* 2010). Finally, a recent review of vascular segmentation algorithms is reported in Moccia *et al* (2018).

Artificial intelligence algorithms

Deep learning with convolutional neural networks (CNNs) have been used to solve numerous complex pattern recognition tasks (Goodfellow *et al* 2016). Deep-learning models use multiple layers of abstraction in a supervised training scheme to learn features, as opposed to hand-crafted feature engineering methods that are used in the papers described above. The investigation of deep CNNs in the recent years has generated tremendous progress on many computer vision tasks in medical image analysis (Litjens *et al* 2017). For instance, CNNs have been shown to achieve outstanding performances for segmentation (Ghafoorian *et al* 2017), localization (de Vos *et al* 2017), cancer diagnosis (Mehrtash *et al* 2017a), quality assessment (Abdi *et al* 2017) and vessel segmentation

(Wang *et al* 2017, Bruno *et al* 2018). We recently reported the first fully 3D CNN based solution to automatically segment individual prostate biopsy needles from T2-weighted MRI (Mehrtash *et al* 2018).

In this work, in order to better support clinicians during GYN-HDR brachytherapy procedures, we propose a fully automatic algorithm to segment multiple catheters in intraoperative MRI. To the best of our knowledge, this is the first time that a fully automatic method has been reported for this task. We report accuracy levels comparable with the semi-automatic algorithm described in Mastmeyer *et al* (2017). The proposed method uses CNN followed by simple post-processing steps consisting of DBSCAN clustering (Ester *et al* 1996) and morphological operation.

Materials and methods

Clinical workflow

All procedures were performed in the advanced multimodality image-guided operating (AMIGO) suite (Tempany *et al* 2015) at Brigham and Women's Hospital in Boston, USA. A Syed-Neblett template with 48 holes arranged in a concentric circles grid, with 10 mm minimum spacing between them, was sutured to the patient's perineum. The obturator inserted in the vaginal canal of the patient through the large middle hole of the template (figure 1), and a collection of needles with plastic sheaths (catheters) was steered through the smaller holes and towards the lesion under MRI guidance. The plastic sheaths or catheters later served as insertion channels for the irradiating spot sources. Periodic intraoperative MRI served to verify if catheters were inserted into the target volume and to ensure that they were not puncturing the bladder, the rectum, or the sigmoid colon. The final MRI was acquired upon insertion of all catheters, and segmentation was performed on those MRI. This workflow was outlined in Viswanathan *et al* (2011) and Kapur *et al* (2012).

Data

The segmentation method was evaluated on MRI scans of 50 gynecologic cancer patients who were treated using a total of 826 HDR brachytherapy catheters in the AMIGO suite of Brigham and Women's Hospital, Boston USA. All the MRIs in this study were acquired intraoperatively and immediately after the catheters were inserted. A 3 Tesla 'Magnetom Verio' scanner was employed (Siemens Healthcare, Erlangen, Germany) and the analyzed images were acquired using the 3D T2/FSE protocol (Siemens SPACE, TR/TE = 3000/160 ms, $0.4 \times 0.4 \times 1.0 \, \text{mm}^3$). The volume field of view was standardized with the center of the obturator in line with the scanner system axis and the template base included in the MR images. Catheters were 1.6 mm (16G) in diameter, with tungsten-alloy stylets or needles inside them. They appeared as tubular signal voids (or dark tubes) in the MRI and the number of catheters in a patient ranged between 10 and 35. According to the clinical protocol, CT images were also available for visual catheter identification.

The image analysis reported here was conducted with approval from the Institutional Review Board (IRB) of Brigham and Women's Hospital. Data were shared with all the co-authors of this study also with explicit IRB approval.

Data annotation

In order to train and test the CNN, reference segmentations for all 826 catheters from 50 patients were manually performed by a radiation physicist (HZ), who is also a trained radiation oncologist, using 3D Slicer software (Fedorov *et al* 2012). These labels represented the ground truth (GT) segmentation. Segmentation was performed on axial slices (figures 2 and 3) starting from the proximal catheter tip, which was identified on the first slice on the cranial side where the catheter appeared. The distal end of each catheter was near the template base. Since a CT scan, with catheters clearly visible, was available for each of these data sets, the expert annotator referred to CT images in cases where there was ambiguity in the location of individual catheters.

Data preprocessing

Magnetic field bias correction was performed on each MR image and the volumes were resampled to a voxel size of $1 \times 1 \times 1 \, \text{mm}^3$. Trilinear interpolation was used for intensity images, nearest neighbor for binary label images. Finally, MRI intensity scale was clipped between the 1st and the 99th percentile and the resulting image histograms were normalized between 0 and 1.

Volume partitioning

MRI volumes were partitioned into 3D tiles, in order to overcome the problem of the limited memory of the available GPUs and to maximize the number of training samples.

A tile was defined as a subset of $60 \times 60 \times 60$ voxels ($60 \times 60 \times 60 \, \text{mm}^3$) within the whole 3D scan. Volume partitioning was on a regular grid and, in general, tiles did not overlap. However, since in some cases the MRI vol-
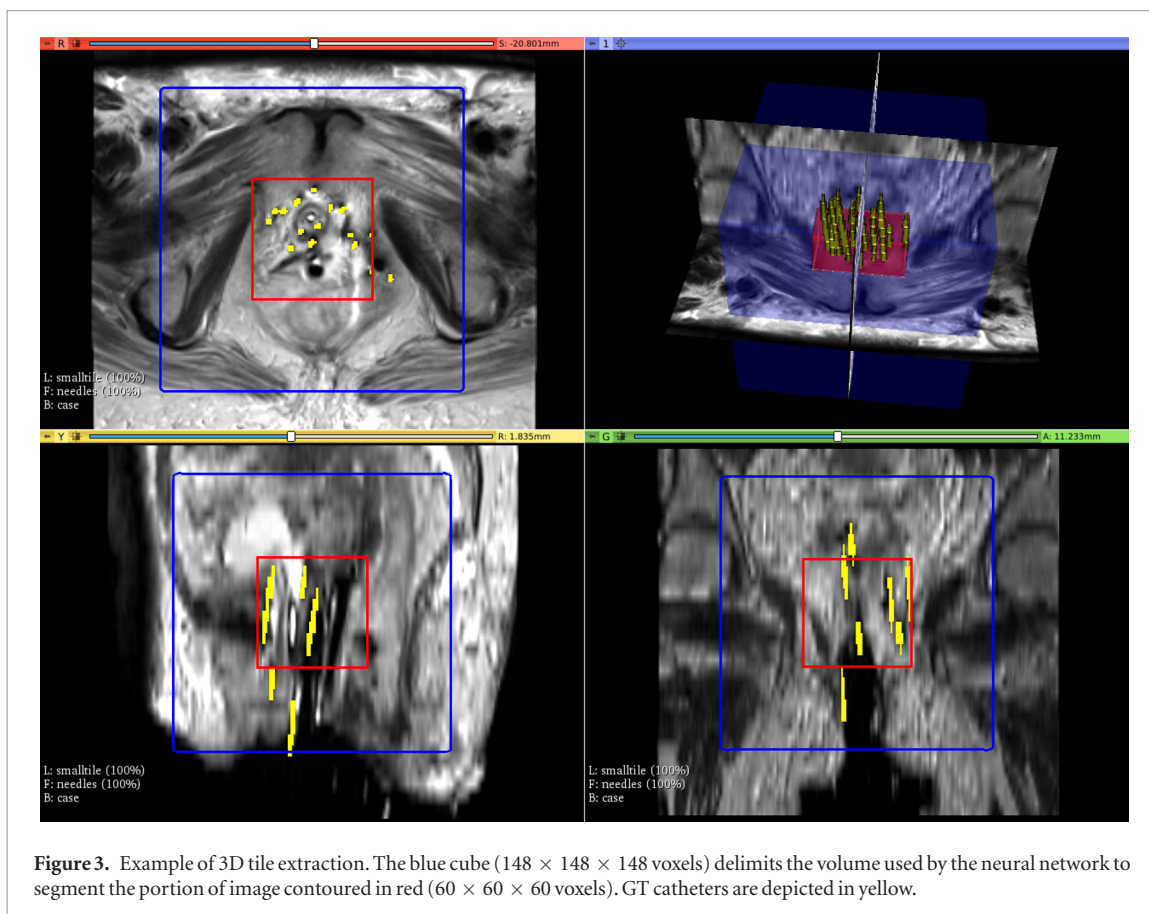
Figure 3. Example of 3D tile extraction. The blue cube ($148 \times 148 \times 148$ voxels) delimits the volume used by the neural network to segment the portion of image contoured in red ($60 \times 60 \times 60$ voxels). GT catheters are depicted in yellow.

ume was not a multiple of tile size, some tiles could partially overlap other tiles. To segment a single patient MRI, all tiles included in the entire volume had to be processed by CNN.

Due to the chosen type of network that provides segmentation by taking into account the neighboring parts of the image, each input tile was generated by including 44 additional voxels per side (padding). In particular, padding was obtained by mirroring the border of the image. As a result, a tile of $148 \times 148 \times 148$ voxels ($148 \times 148 \times 148\,\text{mm}^3$) was fed into the network and a segmented volume of $60 \times 60 \times 60$ voxels was returned. Graphical description of tile extraction is reported in figure 3. In order to allow a consistent evaluation of tiles close to the border, the entire volume was padded by mirroring 44 voxels per side.

Network architecture

The network architecture used was a 3D U-Net (Ronneberger *et al* 2015) consisting of four levels. Each level was composed of two $3 \times 3 \times 3$ convolution layers (with leaky rectify as activation function) followed by a $2 \times 2 \times 2$ max pooling. The number of convolutional kernels increased by going deeper into the U-Net (the amount of kernel was empirically defined) and connections between encoding and decoding sides were present. Dropout (25%) was used only for the first convolution of each level, while batch normalization was used for all of them. Finally, a $1 \times 1 \times 1$ convolution, with sigmoidal activation function, was used to generate the probability map. In figure 4, a schematic network representation is reported.
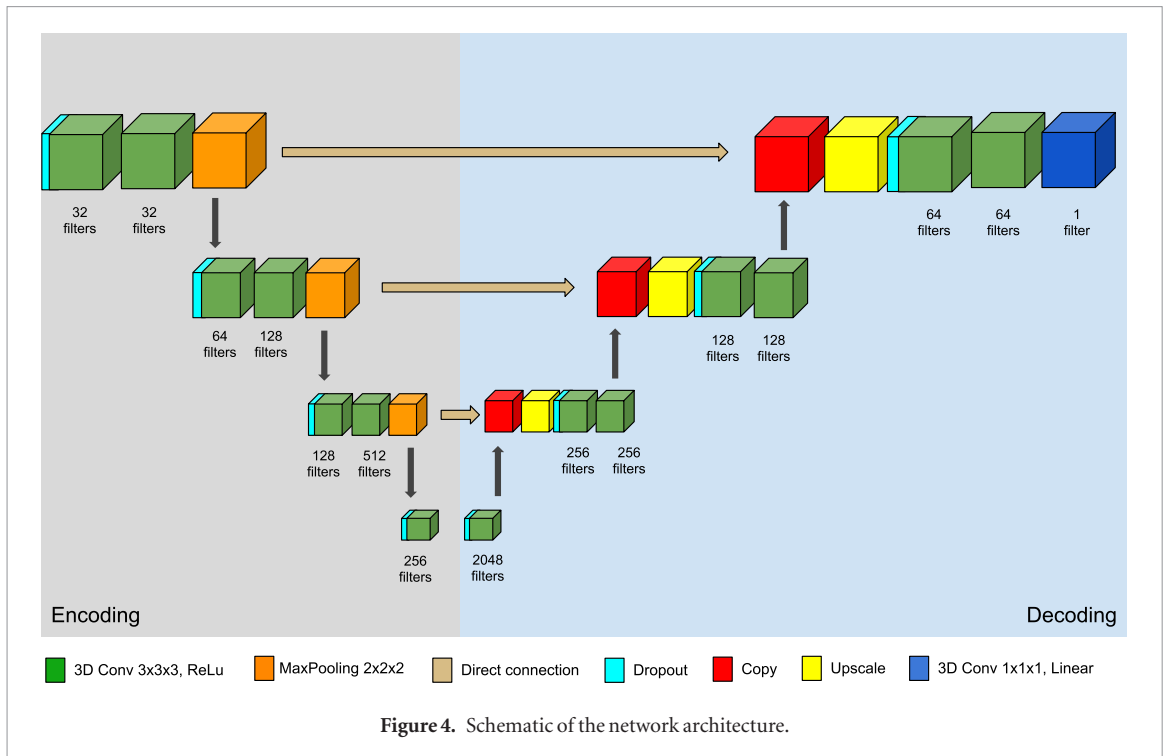
CNNs training

Based on the number of catheters inserted into each of the 50 patients involved in this study, the set of patients was split into 70% for training and 30% for testing. Catheters based splitting was used instead of patient level splitting because of the variable numbers of catheters used in each patient. To improve the learning outcome, the training dataset was augmented by mirroring the axial plane along the left–right direction.

As previously reported, in order to obtain the segmentation of a tile of $60 \times 60 \times 60$ voxels, the U-Net considered also its neighborhood (44 voxels per side) and, as a result, an input tile made of $148 \times 148 \times 148$ voxels ($148 \times 148 \times 148\,\text{mm}$) was fed into the network.

The metrics used to train the classification was Dice Similarity Coefficient (DSC) (Dice 1945). Learning rate was set to 10^{-4} and L2 regularization was used. Batch size was equal to 2.

To train the segmentation network only the tiles containing more than 20 voxels labeled as catheters were used (this threshold was empirically selected). The corresponding manual segmentations performed by the physicist served as GT labels to train the network.

Figure 4. Schematic of the network architecture.

Multiple labels generation and voting

The segmentation network was trained using five different subsets of the training-validation dataset to produce five different models (Mehrtash *et al* 2018). The candidate labels generated by each of the five models were fused together using majority voting. We will refer to this approach as ensemble training. The segmentation generated by majority voting was post-processed by a DBSCAN algorithm that removed clusters located into low density point regions (that is noise obtained as output from CNN). Finally, based on empirical observation that islands smaller than 50 voxels were due to noise, morphological filters were used to remove these small clusters.

Dataset segmentation and evaluation metrics

To obtain robust performance from the network and facilitate a thorough evaluation on the whole 50 patient population, 4-fold cross-validation was performed. This means that four subsets of unique testing patients were separated and the ensemble training was executed each time on the remaining patients.

In figure 5, the k-fold and ensemble procedure are explained by means of pseudo-code and a diagram.

To evaluate the quality of the catheter segmentation, a set of distance metrics between automatic and GT segmentation were quantified (see equations (1)–(6)). Pairwise comparison was performed for each segmented catheter and its GT. After visual inspection of the obtained labels, aiming to guarantee the absence of macroscopic segmentation errors, each automatic catheter was matched to the nearest corresponding GT catheter. The search volume for the matching catheter was obtained by morphologically dilating the investigated GT catheter mask. The number of spurious voxel clusters misclassified as catheters (false positive), as well as the number of missing catheters (false negative), were recorded separately.

If $d(a,b)$ is the euclidean distance between two points and A and B are defined as the non-zero voxels in the reference and the automatic segmentations respectively, Hausdorff distance (HD) (Aspert *et al* 2002) is computed as:

$$\mathrm{HD}(A, \ B) = \max(D(A,B), \ D(B,A)) \tag{1}$$

with D defined as:

$$D(A, \ B) = \max_{a \in A} \min_{b \in B} d(a, b). \tag{2}$$

In order to exclude outlier distance values, 95th percentile of distances (95th$D$) between labels was computed as:

$$95\mathrm{th}D(A,B) \ = \ \frac{95\mathrm{th}d(A,B) \ + \ 95\mathrm{th}d(B,A)}{2} \tag{3}$$

where:

$$95\mathrm{th}d(A,B) \ = \ q_{0.95}(\min_{b \in B} d(a,b)) \ \forall \ a \in A. \tag{4}$$

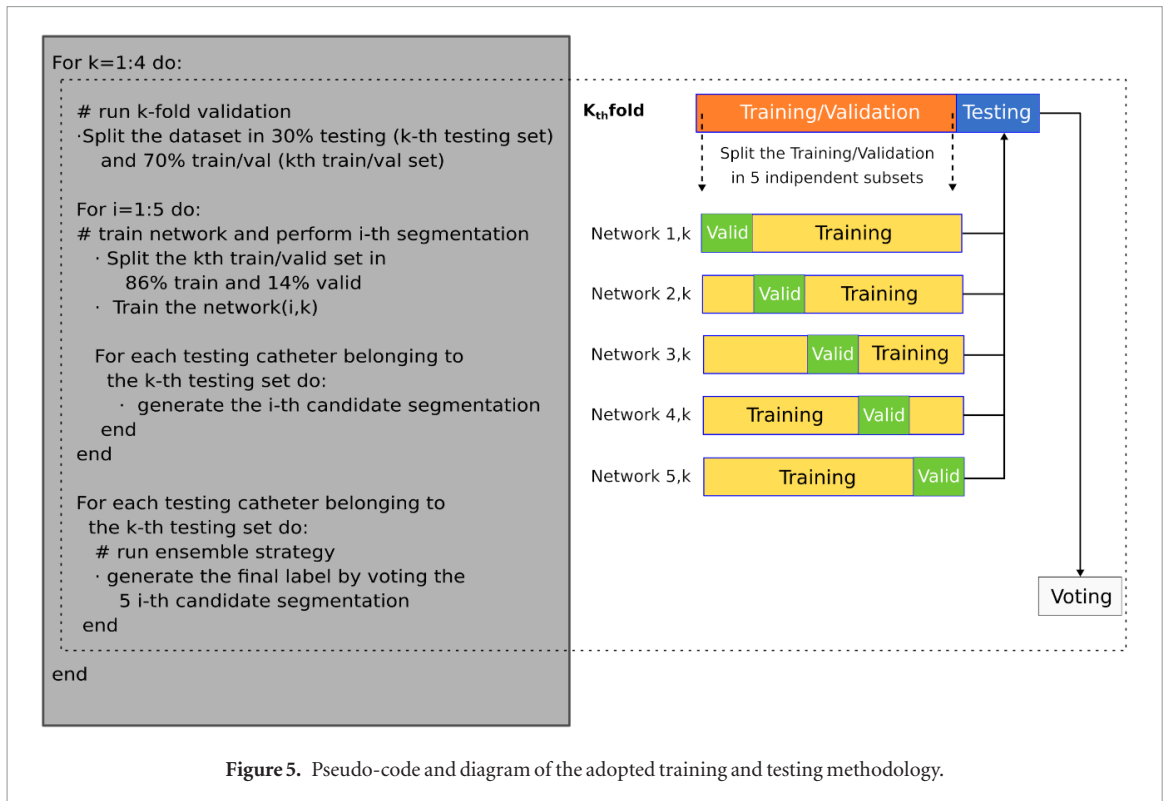Finally, average distance (D_{avg}) was quantified as:

Figure 5. Pseudo-code and diagram of the adopted training and testing methodology.

$$D_{avg}(A,B) \;=\; \frac{d_{avg}(A,B) + d_{avg}(B,A)}{2} \tag{5}$$

with:

$$d_{avg}(A,B) \;=\; \frac{1}{|A|}\sum_{a \in A} \min_{b \in B} d(a,b). \tag{6}$$

$|A|$ is defined as the number of points belonging to A.

The distance metric was chosen because it provides an additional metric that is appropriate for long tubular objects, and often more intuitive for users to understand. For each pair of catheters DSC was also quantified, taking into account that, for small structures (volume smaller than 5 cc), DSC between 0.4 and 0.6 represents a good agreement (Peroni *et al* 2013). All error metrics were computed by using the routines implemented in open source software Plastimatch (Zaffino *et al* 2016).

Implementation and deployment
The reported methods were developed in Python by using SimpleITK (Lowekamp *et al* 2013), NumPy (Oliphant 2006), Theano (Al-Rfou *et al* 2016) and Lasagne (Dieleman *et al* 2015) libraries. The networks were trained on a GNU/Linux workstation equipped with an Intel Xeon CPU, 64 GB of RAM and a Nvidia P6000 GPU with 24 GB of RAM. Finally, to facilitate sharing and deployment of the trained models with other researchers, the 3D Slicer DeepInfer module was used (Mehrtash *et al* 2017b). In addition to the reported ensemble of trained networks, a computationally 'lighter' version is also provided on DeepInfer, which only uses the best network (as reported on training data) from the ensemble. This provides the interested users with one option that may be more accurate and another that is faster.
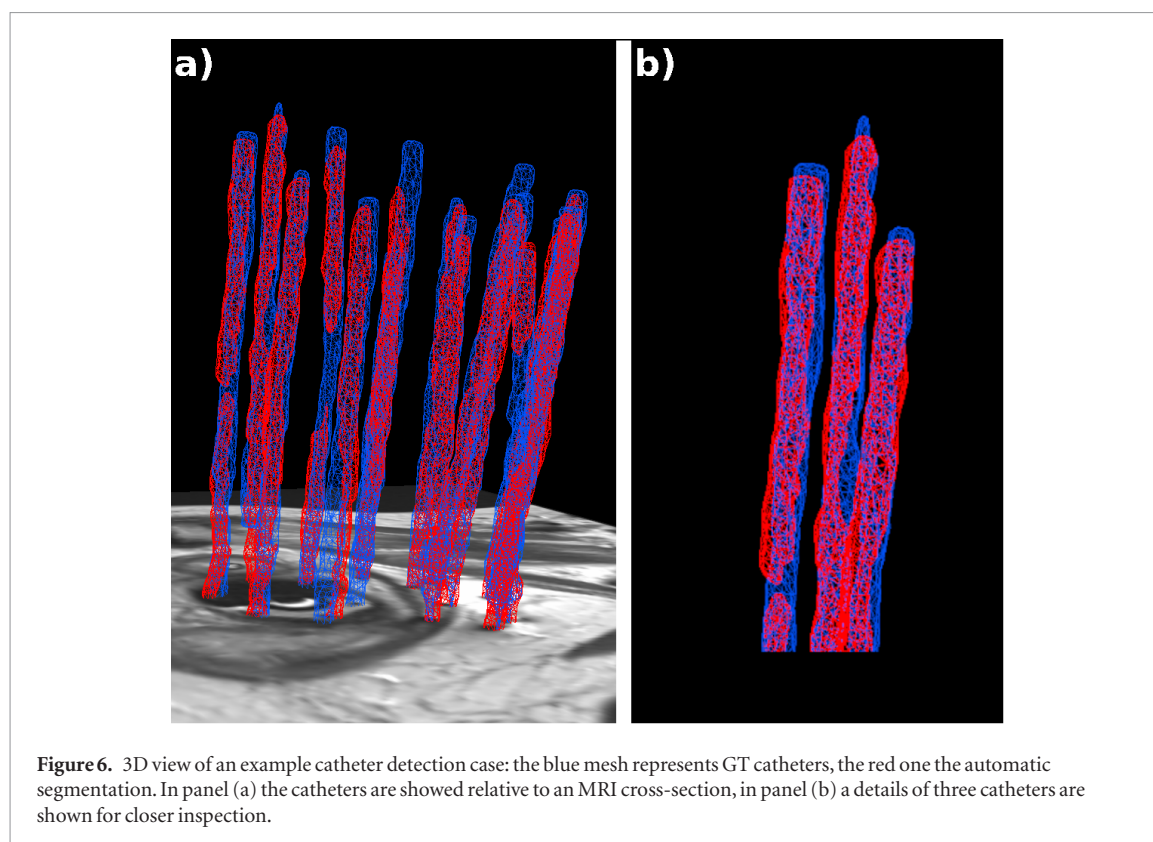
Results

For each pair of GT/automatic catheter, DSC, HD, 95th$D$ and D_{avg} were computed to quantify the accuracy of the final segmentation. Mean and standard deviation values for each metric are reported in table 1.

Overall, the highest distance values were found along the craniocaudal direction (figure 6, top and bottom and gaps of automatic/GT tubes). The number of false positive catheters was 55 out of 826, which corresponds to 6.7%. Undetected GT catheters (false negatives) were 12 out of 826 (1.5%).

It is important to highlight that catheters close to the center of the obturator, which are not well distinguished in the image from the signal void created by the obturator itself, were well identified. An illustrative final segmentation is shown in figure 6.

Table 1. Segmentation accuracy for obtained labels. Dice similarity coefficient, Hausdorff Distance, 95th percentile and average distance between catheters pairs are reported.

	DSC	HD (mm)	95th$D$ (mm)	D_{avg} (mm)
AVG	0.60	15.9	6.9	2.0
STD	0.17	20.5	9.5	3.4

Figure 6. 3D view of an example catheter detection case: the blue mesh represents GT catheters, the red one the automatic segmentation. In panel (a) the catheters are showed relative to an MRI cross-section, in panel (b) a details of three catheters are shown for closer inspection.

Visual inspection by axial views of the best, worst and average cases are reported in supplementary materials as animated gifs.

The maximum number of epochs needed to identify the optimal set of parameters was equal to 48, with a training time for each network of 2 d. Once the weights were computed, the inference time or the time it took to label a test MRI (using ensemble of five models) was 9 ± 2.5 min.

Segmentation accuracy was also quantified for the catheters generated by the faster or 'light' version of the workflow (one model was used, not the ensemble). The same evaluation procedure used for the voted label approach was adopted for the single training labelmaps. The obtained results are reported in table 2.

False negative and false positive catheters were 10 and 114, that correspond to 1.2% and 13.8% respectively. By using the above described single-model based inference, the computational time was 2 ± 0.5 min.

Discussions

In this work we proposed a new deep learning method for supporting radiation physicists during MRI-guided HDR GYN brachytherapy treatment planning using fully automatic segmentation or digitization of brachytherapy catheters on post-insertion MRI scans. The method, based on processing MRI as multiple 3D subvolumes (tiles), uses CNNs.

To the best of our knowledge, this is the first reported method for fully automatic segmentation of catheter collections from gynecological MRI. Mehrtash *et al* (2018) first proved the capability of CNNs to localize a single prostate biopsy needle in intraoperative MRIs. In the gynecologic brachytherapy scenario the task is more challenging, due to the large number of inserted catheters (up to 35 in the cases analyzed in this work) and to the difficulty in precisely defining region of interest an *a priori* region that contains all the catheters in the image.

In more detail, we would like to underscore the intrinsic difficulty of identifying plastic catheter sheaths in MR images. In fact, since no MRI signal is emitted by this type of material, these signal voids appear as black circles in the axial slices and tube segments in coronal and sagittal slices. This leads to two main sources of misrecognition: (1) the possibility to be confused with vessels and other tubular structures (such as urinary catheters) and,

Table 2. Segmentation accuracy of labels generated by the light (non-ensembled) version of the workflow. Dice similarity coefficient, Hausdorff Distance, 95th percentile and average distance between catheters pairs are reported.

	DSC	HD (mm)	95th D (mm)	D_{avg} (mm)
AVG	0.61	12.5	5.4	1.6
STD	0.16	17.2	7.9	2.8

(2) if not completely surrounded by tissues, the difficulty to be distinguished from other signal voids in the MRI that are created by the plastic obturator itself, as well any air in the vaginal canal. The task is even more challenging, since the algorithm was trained and tested on post-procedural MRI acquired in an intraoperative setting, that have lower image quality compared to images obtained in a diagnostic setting when more time is available for the acquisition. To mitigate these issues, it is important to preserve the 3D nature of MRI, and use 3D convolutional filters rather than use a 2D approach. However, this leads to increased computational complexity and processing time. In particular, due to limited GPU memory, in implementing the training solution, trade-offs between small batch size, shallow network and small number of filters had to be carefully balanced.

The choice to execute multiple (here 5) trainings increased robustness regarding the computation of network weights and reduced risk of overfitting to training data.

Obtained segmentations were evaluated by individual corresponding catheters to each other. It is important to note that that this strategy was needed to specifically quantify the accuracy for each catheter and to avoid misleading error estimates by comparing entire point clouds of closely spaced catheters.

In addition to a good accuracy (average distance equal to 2.0 ± 3.4 mm), the proposed methodology does not require any manual interaction from the user. Catheters close to the center of the obturator and partially surrounded by air were also detected, further proving the effectiveness of the proposed workflow. DSC values of 0.60 ± 0.17 were obtained, which are competitive for thin tubular structures. Peroni *et al* (2013), in fact, reports satisfying DSCs of 0.4 and 0.6 for structures smaller than 1 cc and 5 cc respectively. The catheters investigated in this work had an average volume equal to 0.8 ± 0.5 cc (min–max range 0.1–3.1 cc).

In order to mimic realistic scenario of brachytherapy treatment planning, a light workflow was deployed in 3D Slicer via the DeepInfer module, making it usable on a typical desktop computer. The proposed light version consists of inference based on a single model instead of ensembling results from five models. The model with the best validation metric were used. Post-processing step (DBSCAN and small island removal) was kept unchanged. This approach is a reasonable solution to deploy the software to the research community for further discussion. The tradeoff in using inference based on a single model rather than an ensemble was the increase in the number of false positive catheters (13.8%) but reduced computational time ($4\times$ faster).

The weaknesses of the proposed segmentation method are related to (i) the possibility of non-continuous segmentation of the catheters (that leads to higher Hausdorff distance errors) and (ii) the dependence of the currently trained networks on a single MRI scanner.

MRI guided insertion of GYN brachytherapy catheters remains a procedure that is currently only performed at a limited number of hospitals. The segmentation or digitization of catheters from MRI is a technical hindrance to its seamless adoption in the radiation treatment planning workflow. With the increasing number of MRI scanners being installed in departments of radiation oncology around the United States, we hope that larger and more comprehensive datasets acquired across multiple scanners and centers will become available for robust research solutions to this segmentation task. In the spirit of open science, centers all over the world are invited to use, provide feedback to, and contribute to the deployed open source software. To mitigate the impact of limited training data and to reduce the high cost of data labelling, it would be interesting to investigate the generation synthetic data that mimics the properties of MRI, for example using style transfer from CT to MRI (Jin *et al* 2018).

In this study we do not identify catheters individually (that requires further mesh generation and HDR radioactive source channel assignment) but classify the voxels as catheter or non-catheter. Future efforts will use the obtained labelmaps to generate meshes by taking advantage of the dedicated catheter reconstruction algorithm already developed and available in 3D Slicer. The current version of this reconstruction algorithm (Mastmeyer *et al* 2017) is specifically designed to deal with brachytherapy catheters without any raw segmentation working as initialization. An improved version for mesh generation can be implemented by combining raw image data and output from this CNN. Joining these two strategies could lead to more accurate results, to solve the issue related to possibly split segmentations and to better locate each catheter in the physical space (instead of the discrete image space) and assign them to the correct HDR radioactive source channel.

Conclusions

In this study, for the first time, a fully automatic, deep learning based segmentation of multiple closely spaced catheters from MR images was reported for gynecological brachytherapy treatment planning. Accurate

identification was achieved, despite the presence of confounding structures (e.g. blood vessels, plastic obturators). The algorithm is freely usable via the open source software platform 3D Slicer, enabling open science.

Acknowledgments

ORCID iDs

Paolo Zaffino https://orcid.org/0000-0002-0219-0157
Maria Francesca Spadea https://orcid.org/0000-0002-5339-9583

References

Abdi A H *et al* 2017 Correction to 'Automatic quality assessment of echocardiograms using convolutional neural networks: feasibility on the apical four-chamber view' [Jun 17 1221–1230] *IEEE Trans. Med. Imaging* **36** 1992

Aboofazeli M, Abolmaesumi P, Mousavi P and Fichtinger G 2009 A new scheme for curved needle segmentation in three-dimensional ultrasound images *IEEE Int. Symp. on Biomedical Imaging: From Nano to Macro, 2009 (Boston, MA, 28 June–1 July 2009)* pp 1067–70

Al-Rfou R *et al* 2016 Theano: a Python framework for fast computation of mathematical expressions (arXiv:1605.02688)

American Cancer Society 2018 Cancer facts and figures (https://cancerstatisticscenter.cancer.org)

Anderson R *et al* 2018 Interventional Radiation Oncology (IRO): transition of a magnetic resonance simulator to a brachytherapy suite *Brachytherapy* **17** 587–96

Aspert N, Santa-Cruz D and Ebrahimi T 2002 MESH: measuring errors between surfaces using the hausdorff distance *2002 IEEE Int. Conf. on Multimedia and Expo, 2002. Proc.* (IEEE) vol 1, pp 705–8

Beigi P, Rohling R, Salcudean T, Lessoway V A and Ng G C 2015 Needle trajectory and tip localization in real-time 3D ultrasound using a moving stylus *Ultrasound Med. Biol.* **41** 2057–70

Bruno P, Zaffino P, Scaramuzzino S, De Rosa S, Indolfi C, Calimeri F and Spadea M F 2018 Using CNNs for designing and implementing an automatic vascular segmentation method of biomedical images *Int. Conf. of the Italian Association for Artificial Intelligence* (Berlin: Springer) pp 60–70

de Vos B D, Wolterink J M, de Jong P A, Leiner T, Viergever M A and Isgum I 2017 ConvNet-based localization of anatomical structures in 3D medical images *IEEE Trans. Med. Imaging* **36** 1470–81

Dice L R 1945 Measures of the amount of ecologic association between species *Ecology* **26** 297–302

Dieleman S *et al* 2015 *Lasagne: First Release* (Geneva: Zenodo)

Dimopoulos J C, Kirisits C, Petric P, Georg P, Lang S, Berger D and Pötter R 2006 The Vienna applicator for combined intracavitary and interstitial brachytherapy of cervical cancer: clinical feasibility and preliminary results *Int. J. Radiat. Oncol. Biol. Phys.* **66** 83–90

Ester M, Kriegel H P, Sander J and Xu X 1996 A density-based algorithm for discovering clusters in large spatial databases with noise *KDD'96 Proc. of the 2nd Int. Conf. on Knowledge Discovery and Data Mining (Portland, OR, 2–4 August 1996)* pp 226–31 (Menlo Park, CA: AAAI Press)

Fedorov A *et al* 2012 3D Slicer as an image computing platform for the quantitative imaging network *Magn. Reson. Imaging* **30** 1323–41

Fellner C, Pötter R, Knocke T H and Wambersie A 2001 Comparison of radiography-and computed tomography-based treatment planning in cervix cancer in brachytherapy with specific attention to some quality assurance aspects *Radiother. Oncol.* **58** 53–62

Ghafoorian M, Karssemeijer N, Heskes T, Uden I W, Sanchez C I, Litjens G, Leeuw F E, Ginneken B, Marchiori E and Platel B 2017 Location sensitive deep convolutional neural networks for segmentation of white matter hyperintensities *Sci. Rep.* **7** 5110

Goodfellow I, Bengio Y, Courville A and Bengio Y 2016 *Deep Learning* vol 1 (Cambridge, MA: MIT Press)

Görres J, Brehler M, Franke J, Barth K, Vetter S Y, Córdova A, Grützner P A, Meinzer H P, Wolf I and Nabers D 2014 Intraoperative detection and localization of cylindrical implants in cone-beam CT image data *Int. J. Comput. Assist. Radiol. Surg.* **9** 1045–57

Grover S, Harkenrider M M, Cho L P, Erickson B, Small C, Small W Jr and Viswanathan A N 2016 Image guided cervical brachytherapy: 2014 survey of the American Brachytherapy Society *Int. J. Radiat. Oncol. Biol. Phys.* **94** 598–604

Haie-Meder C *et al* 2005 Recommendations from gynaecological (GYN) GEC-ESTRO Working Group (I): concepts and terms in 3D image based 3D treatment planning in cervix cancer brachytherapy with emphasis on MRI assessment of GTV and CTV *Radiother. Oncol.* **74** 235–45

Hao X, Zygmunt K, Whitaker R T and Fletcher P T 2014 Improved segmentation of white matter tracts with adaptive Riemannian metrics *Med. Image Anal.* **18** 161–75

Hrinivich W T, Hoover D A, Surry K, Edirisinghe C, Montreuil J, D'Souza D, Fenster A and Wong E 2017 Simultaneous automatic segmentation of multiple needles using 3D ultrasound for high-dose-rate prostate brachytherapy *Med. Phys.* **44** 1234–45

Jin C B, Jung W, Joo S, Park E, Saem A Y, Han I H, Lee J I and Cui X 2018 Deep CT to MR synthesis using paired and unpaired data (arXiv:1805.10790)

Kamran S C, Manuel M M, Cho L P, Damato A L, Schmidt E J, Tempany C, Cormack R A and Viswanathan A N 2017 Comparison of outcomes for MR-guided versus CT-guided high-dose-rate interstitial brachytherapy in women with locally advanced carcinoma of the cervix *Gynecol. Oncol.* **145** 284–90

Kapur T, Egger J, Damato A, Schmidt E J and Viswanathan A N 2012 3 T MR-guided brachytherapy for gynecologic malignancies *Magn. Reson. Imaging* **30** 1279–90

Liskowski P and Krawiec K 2016 Segmenting retinal blood vessels with deep neural networks *IEEE Trans. Med. Imaging* **35** 2369–80

Litjens G, Kooi T, Bejnordi B E, Setio A A A, Ciompi F, Ghafoorian M, Van Der Laak J A, Van Ginneken B and Sánchez C I 2017 A survey on deep learning in medical image analysis *Med. Image Anal.* **42** 60–88

Lowekamp B C, Chen D T, Ibáñez L and Blezek D 2013 The design of SimpleITK *Frontiers Neuroinform.* **7** 45

Mastmeyer A, Fortmeier D and Handels H 2016 Efficient patient modeling for visuo-haptic VR simulation using a generic patient atlas *Comput. Methods Programs Biomed.* **132** 161–75

Mastmeyer A, Pernelle G, Ma R, Barber L and Kapur T 2017 Accurate model-based segmentation of gynecologic brachytherapy catheter collections in MRI-images *Med. Image Anal.* **42** 173–88

Mehrtash A, Pesteie M, Hetherington J, Behringer P A, Kapur T, Wells W M, Rohling R, Fedorov A and Abolmaesumi P 2017b Deepinfer: open-source deep learning deployment toolkit for image-guided therapy *Proc. SPIE* **10135** 101351K

Mehrtash A, Sedghi A, Ghafoorian M, Taghipour M, Tempany C M, Wells W M, Kapur T, Mousavi P, Abolmaesumi P and Fedorov A 2017a Classification of clinical significance of MRI prostate findings using 3D convolutional neural networks *Proc. SPIE* **10134** 101342A

Mehrtash A *et al* 2018 Automatic needle segmentation and localization in MRI with 3D convolutional neural networks: application to MRI-targeted prostate biopsy *IEEE Trans. Med. Imaging* **38** 1026–36

Moccia S, De Momi E, El Hadji S and Mattos L S 2018 Blood vessel segmentation algorithms—review of methods, datasets and evaluation metrics *Comput. Methods Programs Biomed.* **158** 71–91

Nguyen H G, Fouard C and Troccaz J 2015 Segmentation, separation and pose estimation of prostate brachytherapy seeds in CT images *IEEE Trans. Biomed. Eng.* **62** 2012–24

Novotny P M, Stoll J A, Vasilyev N V, Pedro J, Dupont P E, Zickler T E and Howe R D 2007 GPU based real-time instrument tracking with three-dimensional ultrasound *Med. Image Anal.* **11** 458–64

O'Donnell L J and Westin C F 2007 Automatic tractography segmentation using a high-dimensional white matter atlas *IEEE Trans. Med. Imaging* **26** 1562–75

Oliphant T E 2006 *A Guide to NumPy* (USA: Trelgol Publishing)

Pernelle G *et al* 2013 Validation of catheter segmentation for MR-guided gynecologic cancer brachytherapy *Int. Conf. on Medical Image Computing and Computer-Assisted Intervention* (Berlin: Springer) pp 380–7

Peroni M, Spadea M F, Riboldi M, Falcone S, Vaccaro C, Sharp G C and Baroni G 2013 Validation of automatic contour propagation for 4D treatment planning using multiple metrics *Technol. Cancer Res. Treat.* **12** 501–10

Pötter R *et al* 2006 Recommendations from gynaecological (GYN) GEC ESTRO working group (II): concepts and terms in 3D image-based treatment planning in cervix cancer brachytherapy—3D dose volume parameters and aspects of 3D image-based anatomy, radiation physics, radiobiology *Radiother. Oncol.* **78** 67–77

Ronneberger O, Fischer P and Brox T 2015 U-net: convolutional networks for biomedical image segmentation *Int. Conf. on Medical Image Computing and Computer-Assisted Intervention* (Berlin: Springer) pp 234–41

Sultana S, Blatt J E, Gilles B, Rashid T and Audette M A 2017 MRI-based medial axis extraction and boundary segmentation of cranial nerves through discrete deformable 3d contour and surface models *IEEE Trans. Med. Imaging* **36** 1711–21

Tempany C M, Jayender J, Kapur T, Bueno R, Golby A, Agar N and Jolesz F A 2015 Multimodal imaging for improved diagnosis and treatment of cancers *Cancer* **121** 817–27

Uherčík M, Kybic J, Liebgott H and Cachard C 2010 Model fitting using ransac for surgical tool localization in 3D ultrasound images *IEEE Trans. Biomed. Eng.* **57** 1907–16

Viswanathan A N, Cormack R, Holloway C L, Tanaka C, O'Farrell D, Devlin P M and Tempany C 2006 Magnetic resonance–guided interstitial therapy for vaginal recurrence of endometrial cancer *Int. J. Radiat. Oncol. Biol. Phys.* **66** 91–9

Viswanathan A N, Kirisits C, Erickson B E and Pötter R 2011 *Gynecologic Radiation Therapy—Novel Approaches to Image-Guidance and Management* (Berlin: Springer)

Viswanathan A N, Szymonifka J, Tempany-Afdhal C M, O'Farrell D A and Cormack R A 2013 A prospective trial of real-time magnetic resonance–guided catheter placement in interstitial gynecologic brachytherapy *Brachytherapy* **12** 240–7

Viswanathan A N, Thomadsen B and American Brachytherapy Society Cervical Cancer Recommendations Committee 2012 American Brachytherapy Society consensus guidelines for locally advanced carcinoma of the cervix. Part I: general principles *Brachytherapy* **11** 33–46

Wang J, Ding H, Bidgoli F A, Zhou B, Iribarren C, Molloi S and Baldi P 2017 Detecting cardiovascular disease from mammograms with deep learning *IEEE Trans. Med. Imaging* **36** 1172–81

Wink O, Niessen W J and Viergever M A 2004 Multiscale vessel tracking *IEEE Trans. Med. Imaging* **23** 130–3

Zaffino P, Raudaschl P, Fritscher K, Sharp G C and Spadea M F 2016 Plastimatch MABS, an open source tool for automatic image segmentation *Med. Phys.* **43** 5155–60