Contour Propagation in CT Scans with Convolutional Neural Networks

Jean Léger*, Eliott Brion, Umair Javaid, John Lee, Christophe De Vleeschouwer, and Benoit Macq

Université catholique de Louvain, Louvain-la-Neuve, Belgium
{jean.leger,eliott.brion}@uclouvain.be

Abstract. Although deep convolutional neural networks (CNNs) have outperformed state-of-the-art in many medical image segmentation tasks, deep network architectures generally fail in exploiting common sense prior to drive the segmentation. In particular, the availability of a segmented (source) image observed in a CT slice that is adjacent to the slice to be segmented (or target image) has not been considered to improve the deep models segmentation accuracy. In this paper, we investigate a CNN architecture that maps a joint input, composed of the target image and the source segmentation, to a target segmentation. We observe that our solution succeeds in taking advantage of the source segmentation when it is sufficiently close to the target segmentation, without being penalized when the source is far from the target.

1 Introduction

Computed tomography (CT) image segmentation is highly useful for a large number of medical applications such as image analysis in radiology [1], treatment planning and patient follow-up in radiotherapy [2] or therapeutic response prediction through radiomics [3]. Classical clinically used segmentation methods include atlas-based methods [4] or active contours [5], but are generally failing in cases where the regions of interest are subjected to large deformations or matter income/outcome. Statistical shape models [6] allow to capture shape variations but require landmarks definition. Pixel-wise classification has also been performed with random forest and SVM classifiers [7], [8], potentially completed by structure enhancing methods such as conditional random fields [9] or graph cuts. However, those methods require a tedious and subjective definition of features. Alternatively, numerous hybrid methods have proposed to combine several complementary segmentation approaches [10], [11], [12]. Nevertheless, none of them closes the challenging task of CT image segmentation.

In parallel, the recent advances in computing capabilities and the availability of representative datasets allowed deep learning approaches to reach impressive segmentation performance, competing with state-of-the-art segmentation tools in many fields of medical imaging [13]. The main advantage of deep learning with

* The first two authors contributed equally.

respect to the aforementioned approaches is its high versatility and its capability to learn automatically a complex input-output mapping through the tuning of a large number of inner parameters. Fully convolutional neural networks [14] gained in popularity compared to patch-based approaches since they benefit from a higher computational efficiency and a better integration of contextual information. In particular, the u-net fully convolutional neural network architecture [15] resulted in good segmentation performances on 2D medical images, as well as its 3D extension [16] and 3D V-Net variation [17].

Deep learning segmentation algorithms have been successfully used to segment structures on CT images in the head and neck [18], the pelvic [19] and the abdominal [20], [21] regions. Those deep models rely on the assumption that the learned filters are able to identify automatically, from the training data, the features that are relevant for the task at hand. However, in many applications, additional information, also named *prior information*, is available and should be exploited to constraint the segmentation process towards the desired output. Recently, shape prior constraints have been incorporated in deep neural networks through PCA [22], resulting in an improved segmentation robustness and accuracy. Anatomical constraints have also been accounted for through the combination of several deep neural networks [23].

Similarly, the availability of a coarse or approximate segmentation may provide a valuable prior information to guide a fine-grained segmentation process. Examples of practical applicative scenarios in which such approximate segmentation is available include dense 3D segmentation (for which only a subset of the slices have been segmented), but also the adaptation of radiotherapy treatments (for which the segmentation of an initial image should be propagated to subsequent images aquired along the treatment).

In this work, an approach has been developed to segment 2D CT image slices of the bladder using the manual segmentation (drawn by an experienced radiation oncologist) on an adjacent slice as prior knowledge. Our goal is to design a deep network architecture able to capture both the bladder appearance statistics and the deformation occurring between adjacent slices. In that sense, it allows to *propagate* the segmentation between adjacent slices. To demonstrate the relevance of our approach, we analyze quantitatively how the segmentation result improves with the proximity of the adjacent slice. Note that slices segmentation of the bladder is not a trivial task due to the high shape and size variation of the bladder, the presence or not of contrast material in it and the poor contrast between its wall and surrounding soft tissues.

Section 2 introduces the CT images that have been considered in our study and their preprocessing. Section 3 explains how to turn the segmentation problem into a learning problem, by defining pairs of input/output training samples based on the CT images and their manual annotations by medical doctors. The network architecture is also given in the same section. In Sect. 4, the results are provided and discussed. Conclusions are given in Sect. 5.

2 Materials and Preprocessing

Our data consist of 3D CT volumes (planning CTs, also named PCTs) from patients with prostate cancer. Each patient underwent External Beam Radiation Therapy (EBRT) at CHU-UCL-Namur (198 patients) or at CHU-Charleroi – Hôpital A. Vésale (141 patients). The images coming from both hospitals are shuffled in our dataset and images with and without contrast agent are present in it. For each patient, the bladder was manually delineated in the PCT by an experienced radiation oncologist. This has been considered as the ground truth in this work. The use of these retrospective, anonymized data for this study as been approved by each hospital ethics committee. The data were pre-processed according to the following steps: data re-sampling, slice selection and cropping and intensity range thresholding.

The 3D CT volumes (stacks of 2D CT slices) extracted in the hospitals under DICOM format have a different pixel spacing and slice thickness across different patients. In order to ensure data uniformity over the entire dataset, all the 3D CT volumes have been re-sampled through a linear interpolation according to a 1x1x1 mm regular grid. The bladder contours drawn by the clinicians are stored as 3D binary masks with same sampling as the CT volumes. The DICOM format manipulation and re-sampling are performed using the opensource platform OpenReggui.[1]

For every patient, a single 2D axial slice where the bladder is present is randomly selected, together with several adjacent slices, from the 3D CT volume. This is further discussed in Sect. 3.1. Every selected slice is cropped to 192x192 pixels tile centered on the bladder. The reason for cropping is that it allows faster experimentation. Our investigations (not described in this paper) have shown that running the algorithm on the entire slice instead of an image tile centered on the region of interest leads to comparable segmentation accuracy.

All pixels intensities above 3000 HU (Hounsfield unit) and below -1000 HU are respectively set to 3000 HU and -1000 HU to avoid artifacts in the high Hounsfield unit range. The lower and upper bounds are respectively determined from the Hounsfield values of air and cortical bone.

3 Formulating the Contour Propagation as a Learning Problem

Our goal is to segment target 2D CT image slices of the bladder using the manual delineation on an adjacent slice as prior knowledge. In this section, the target and adjacent slices selection is firstly explained. Then, we present the CNN architecture used for the segmentation and how the prior information is included in it.

[1] `https://openreggui.org/`.

3.1 Prior Definition and Computation

Each 3D CT volume is made of a stack of 2D slices along the axial axis. For every patient, the bladder only appears on a set of slices indexed by $p \in [0, 1, ..., P-1]$ where the bladder top and bottom slices are respectively indexed by $p = 0$ and $p = P - 1$, with P the number of slices containing the bladder. For every patient, a single *target* slice indexed by p_t has been randomly selected such that $p_t \in [0, (P-1) - 10]$.

Then, for every target slice indexed by $p = p_t$, the corresponding prior knowledge is extracted from an *adjacent* slice indexed by $p_q = p_t + q$ with $q \in [0, 1, ..., 10]$. Hence, the indices p_q of the prior slices are always such that $p_q \in [0, (P-1)]$. The prior knowledge used in this work is the manual delineation (contour) drawn by the clinician on the qth adjacent slice. This contour is represented as a binary mask $\mathbf{Y}^{(k)[q]} \in \{0,1\}^{M \times M}$ defined as

$$Y_{m,n}^{(k)[q]} = \begin{cases} 1 & \text{if } (m,n) \text{ is in the bladder} \\ 0 & \text{otherwise} \end{cases} \tag{1}$$

where (m, n) is the pixel position within the binary mask $\mathbf{Y}^{(k)[q]}$, k is the index of the binary mask in the dataset and M is the image tile width.

3.2 Network Architecture and Learning Strategy

The u-net fully convolutional neural network has been used for the bladder segmentation. The architecture of the network is shown in Fig. 1 with the hyper-parameters used in this work. The network input is expanded from one to two channels. The additional channel is used to enter the prior knowledge into the network.

More precisely, let us assume that the network input for the kth data sample is denoted by $\mathbf{X}^{(k)} \in \mathbb{R}^{M \times M \times C}$, where C is the number of input channels. The corresponding labels are a binary mask $\mathbf{Y}^{(k)} \in \{0,1\}^{M \times M}$ equal to one within the bladder and equal to zero everywhere else. The notation $\mathbf{X}_{:,:,c}^{(k)}$ is used in order to denote the cth network input channel. The input is the concatenation of the target image $\mathbf{I}^{(k)} \in \mathbb{R}^{M \times M}$ with the prior mask given on the qth adjacent slice above the target slice: $\mathbf{X}_{:,:,1}^{(k)} = \mathbf{I}^{(k)}$ and $\mathbf{X}_{:,:,2}^{(k)} = \mathbf{Y}^{(k)[q]}$ as shown in Fig. 2. Eleven different networks are trained, each for a different distance between the target slice and the adjacent slice. Formally, for $q \in [0, 1, ..., 10]$, eleven different models $\mathcal{M}_q$ are trained, each of them with a training set $\mathcal{S}_{train}^{[q]}$ given by $\mathcal{S}_{train}^{[q]} = \left\{ \left(\mathbf{X}^{(j)}, \mathbf{Y}^{(j)}\right) \,\middle|\, j \in [0, 1, ..., J_{train} - 1], \mathbf{X}_{:,:,1}^{(j)} = \mathbf{I}^{(j)} \text{ and } \mathbf{X}_{:,:,2}^{(j)} = \mathbf{Y}^{(j)[q]} \right\}$, where J_{train} is the number of target tiles in the training set.

The network architecture, inspired from [15], is illustrated in Fig. 1. It takes as input either one channel (if the CNN is used without prior knowledge, see Sect. 4.1) or two channels (if the CNN is used with prior knowledge) and outputs a prediction for the target tile bladder segmentation. The input goes through

a contracting path (left side) to capture context and an expanding path (right side) to enable precise localization.

In the contracting path, a collection of hierarchical features are learned thanks to successive convolutions and max-pooling operations. Successively, two series of 3x3 convolutions, followed by a ReLu activation and batch normalization are applied before applying a max-pooling 2x2. After each max-pooling step, the number of feature maps is doubled in order to allow the network to learn many high level features. From the features learned in the contracting path, the expanding path increases the resolution via successive 2x2 transposed convolutions (i.e. a 2x2 up-sampling followed by a 2x2 convolution), 3x3 convolutions and ReLu activations in order to recover the original tile size.

In the last layer, a sigmoid is applied and the network outputs the probability for each pixel to belong to the bladder. To obtain the final segmentation, a threshold of 0.5 is chosen. Hence, every pixels with a score above the threshold is predicted to belong to the bladder, while other pixels are predicted not to belong to the bladder.

The network is trained with the Dice loss. For the jth training example, the Dice loss between the predicted segmentation $\hat{\mathbf{Y}}^{(j)} \in [0,1]^{M \times M}$ and the reference segmentation $\mathbf{Y}^{(j)} \in \{0,1\}^{M \times M}$ is given by

$$\mathcal{L}\left(\hat{\mathbf{Y}}^{(j)}, \mathbf{Y}^{(j)}\right) = -\frac{2\sum_{m=0}^{M-1}\sum_{n=0}^{M-1} Y_{m,n}^{(j)} \hat{Y}_{m,n}^{(j)}}{\sum_{m=0}^{M-1}\sum_{n=0}^{M-1}\left(Y_{m,n}^{(j)} + \hat{Y}_{m,n}^{(j)}\right)}. \tag{2}$$

The fact that each training tile has at least one pixel belonging to the bladder ensures that the denominator is non-zero. This loss drives the output segmentation to have a large overlap with the ground truth segmentation. This network is implemented in Keras, with TensorFlow back-end. The optimization algorithm used is Adam with a learning rate of $1e-4$, a batch size of 20 and the model is trained for 200 epochs. The other parameters (such as initialization strategy and learning rate decay) are left to Keras default. All the tiles in the dataset are shifted and scaled by the mean and the standard deviation of the tiles intensities over the training set. We use the validation set to earlystop the training. Training data is augmented using rotation, shift, shear, zoom and horizontal flip.

The target slices and their set of adjacent slices are randomly assigned to the training set (179 patients), the validation set (80 patients) and the test set (80 patients). Note that the patient distribution across the training, validation and test sets is the same for the eleven different models.

4 Results and Discussion

In this section, the validation metric and the comparison baselines are first presented. Then, our results are discussed.

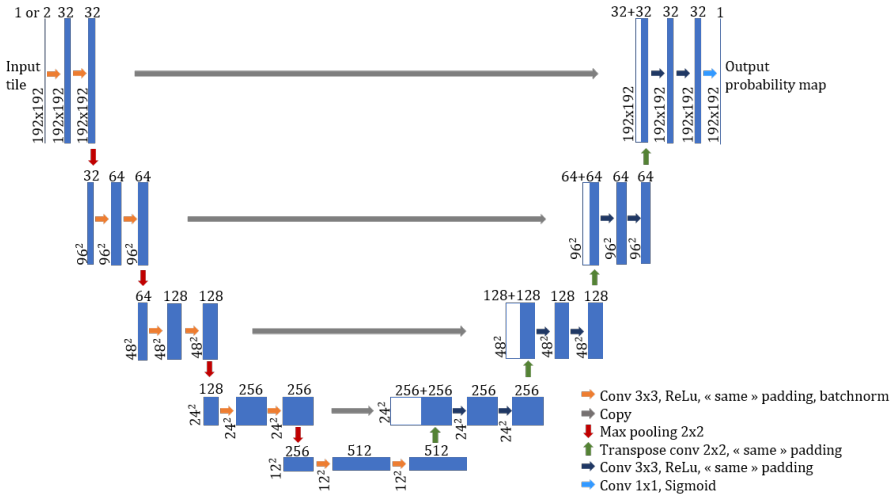

Fig. 1. The network architecture used is the u-net. Each blue rectangle represents the feature maps resulting from a convolution operation, while white rectangles represent copied featured maps. For the convolutions, the zero padding is chosen such that the image size is preserved ("same" padding). This figure is adapted from [15].

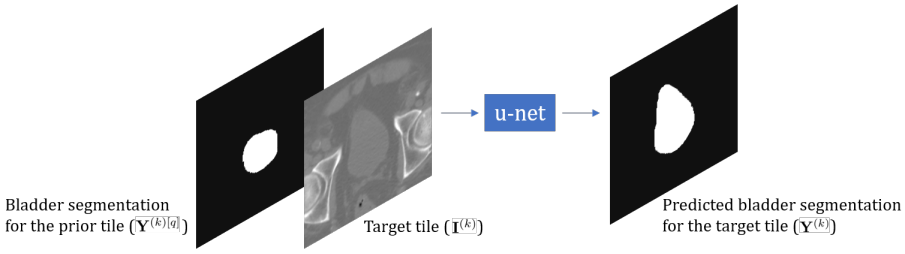

Fig. 2. The proposed u-net network takes as input the bladder segmentation for the prior tile and the target tile intensities. It outputs a prediction of the bladder segmentation for the target tile.

4.1 Validation Metrics and Comparison Baselines

In order to evaluate our results, we use three metrics: the Dice similarity coefficient (DSC) and the Jaccard index (JI) measure the overlap between two binary masks, while the Hausdorff distance (HD) assesses the distance between the contours (i.e. the sets of points located at the boundary of the binary masks) extracted from those binary masks. Most results are discussed based on the DSC since it is quite popular in the literature.

The DSC and the JI both reach one in case of perfect overlap and zero if there is no overlap between both binary masks. The DSC is slightly different from the Dice loss used for the network training. Indeed, the DSC is computed between two binary images, the ground truth and the thresholded network prediction, whereas the Dice loss was computed between the ground truth and the network prediction to allow the backward propagation. The JI is defined as the intersection of both binary images over their union. The HD is computed between the ground truth contour $\mathcal{C}_1$ and the prediction contour $\mathcal{C}_2$. More precisely, the HD is the greatest of all the distances from a point in one contour (both $\mathcal{C}_1$ and $\mathcal{C}_2$ need to be considered) to the closest point in the other contour.

The prediction performed by our network is compared to three different baselines. The *copy* baseline uses the binary mask $\mathbf{Y}^{(k)[q]}$ as a valid approximation of $\mathbf{Y}^{(k)}$. Since the binary mask $\mathbf{Y}^{(k)[q]}$ is given as an input to the network, our algorithm should ideally outperform this baseline.

The u-net network has also been trained without prior knowledge. The network input is simply the target image and has a single channel such that $\mathbf{X}^{(k)}_{:,:,1} = \mathbf{I}^{(k)}$. This is shown in Fig. 1 and denoted as the *CNN without prior* baseline.

The results obtained with our CNN have been compared to the classical registration-based contour propagation. It has been denoted as the *registration* baseline. Such an approach computes a deformation field from the target image $\mathbf{I}^{(k)}$ (fixed image) to the image on the qth adjacent slice $\mathbf{I}^{(k)[q]}$ (the moving image). Applying the inverse of this deformation field to the moving image allows to align the moving image with the fixed image. In turn, applying the inverse of the deformation field to the binary mask on the adjacent slice $\mathbf{Y}^{(k)[q]}$ yields an estimation of the target binary mask. The registration has been performed using SimpleElastix.[2] The registration components have been chosen according to the recommendations of the Elastix's user manual [24]. In particular, the mutual information metric, the B-splines transformation and the adaptive stochastic gradient descent optimizer have been used together with a multi-resolution strategy.

4.2 Discussion

In Fig. 3, the DSC and HD between the output segmentation of the CNN with prior and the ground truth segmentation are computed and averaged over the test set for several distances (i.e. from 1 mm to 10 mm) between the target and adjacent slices. This distance is named the inter-slices distance below. The average DSC and HD obtained with the CNN with prior is also compared to the results obtained with the three aforementioned baselines. Table 1 provides the mean and standard deviation of the DSC and JI for all the considered methods and an inter-slices distance equal to 2, 4 and 8 mm. The following observations can be done based on Fig. 3 and Table 1.

– For a large inter-slices distance (i.e. $8 \leq q$), the CNN with prior does not clearly outperform the CNN without prior. This means that the CNN with

[2] `http://simpleelastix.github.io`.

prior struggles to improve the target segmentation when the prior is not sufficiently correlated with the target segmentation. However, the CNN with prior clearly outperforms the result of the registration-based approach in this region. This is illustrated in the first column of Fig. 4.
- For an intermediate inter-slices distance (i.e. $4 \leq q < 8$), the CNN with prior slightly outperforms the CNN without prior. However, the DSC of the copy is still low in this interval. This means that the CNN with prior is able to exploit jointly the information present in the target image and the prior mask. Furthermore, the prior exploitation increases on average as the prior mask becomes more relevant (i.e. for decreasing q). This is illustrated in the second column of Fig. 4.
- For a small inter-slices distance (i.e. $q < 4$), the CNN with prior performs like the registration-based approach, which is known to work efficiently when the moving (adjacent) and the fixed (target) images are close to each other. This means that the CNN with prior is able to capture information in the prior mask when it is relevant for the segmentation task. This is illustrated in the third column of Fig. 4.

From those observations, we can conclude that the CNN with prior reaches or exceeds the segmentation performance of alternative approaches for all the considered inter-slices distances. Furthermore, the prior knowledge is more intensively exploited as its correlation with the ground truth increases.

From Table 1, it appears that the variance of the considered metrics is relatively high compared to the difference in their mean computed with the different baselines. A finer analysis of the statistical distribution of our segmentation results reveals that, even if the average performance is only moderately better with prior than without prior, the exploitation of the segmentation mask in an adjacent slice (i.e. the prior) prevents a complete failure of the segmentation in worst cases. Indeed, the DSC is always larger than 0.78 in presence of prior, while being lower than 0.8 for about 15 % of the images in absence of prior, with worst cases below 0.6 and reaching 0.2. Hence, we conclude that our proposed framework has a limited impact on the mean DSC performance but significantly increases the robustness of the approach, which is certainly important in practice (e.g. regarding organ segmentation for radiotherapy treatment).

Figure 3 and Table 1 also include the segmentation performance of a single CNN (in opposition to the eleven CNNs) trained over all the training samples, hence corresponding to different inter-slices distances. This CNN is named *CNN with prior single*. This unique CNN is then tested on the different inter-slices separately. It appears that this network improves the segmentation performance with respect to the CNN without prior for all the considered inter-slices distances, even if it is outperformed by the CNNs trained for a specific inter-slide distance for $q < 4$. This result is encouraging for applications where we cannot define a proper inter-slices distance (e.g. regarding the propagation of a contour over a temporal sequence of images).

The training time was respectively 446 s and 488 s for the CNN without prior and for the CNN with prior. This training time is an average over the training

times obtained for the three inter-slices distances considered in Table 1. The inference time was respectively ~ 0 s, 3.58 s, 0.057 s and 0.184 s for the copy, the registration, the CNN without prior and the CNN with prior. This inference time is the time necessary to predict the segmentation of a single image, averaged over the three inter-slices distances and over all the test images. It appears that the CNN with and without prior have comparable training times. It can be also observed that the inference time of CNNs is one to two orders of magnitude below the inference time of the registration-based method. The CNNs have been trained and tested with nVidia cuDNN acceleration on a computer equipped with a GPU nVidia GeForce GTX 1080 Ti 11Gb.[3]

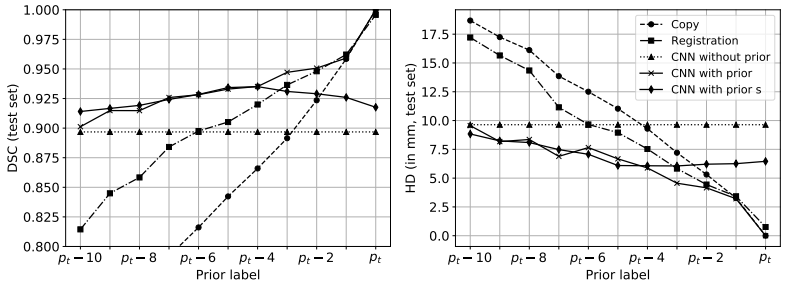

Fig. 3. Influence of the prior on the DSC and on the HD for test set examples. DSC: Dice similarity coefficient. HD: Hausdorff distance.

Table 1. Comparison of the different segmentation algorithms regarding to their DSC and JI on test set. See text for details. DSC: Dice similarity coefficient. JI: Jaccard index. CNN wop: CNN without prior, CNN wp: CNN with prior (eleven different networks are trained, each one for a given inter-slice distance), CNN wps : CNN with prior single (one single network is trained on all different inter-slice distances).

	$p_t - 8$		$p_t - 4$		$p_t - 2$	
Algorithm	DSC	JI	DSC	JI	DSC	JI
Copy	.761±.176	.645±.214	.866±.122	.781±.166	.924±.077	.867±.120
Registration	.858±.147	.775±.186	.920±.101	.864±.130	.948±.048	.905±.080
CNN wop	.897±.159	.840±.191	.897±.159	.840±.191	.897±.159	.840±.191
CNN wp	.915±.103	.856±.140	**.935±.071**	**.885±.104**	**.951±.036**	**.908±.062**
CNN wps	**.919±.088**	**.861±.128**	**.935±.055**	.883±.088	.929±.063	.873±.099

[3] `https://developer.nvidia.com/cudnn`.

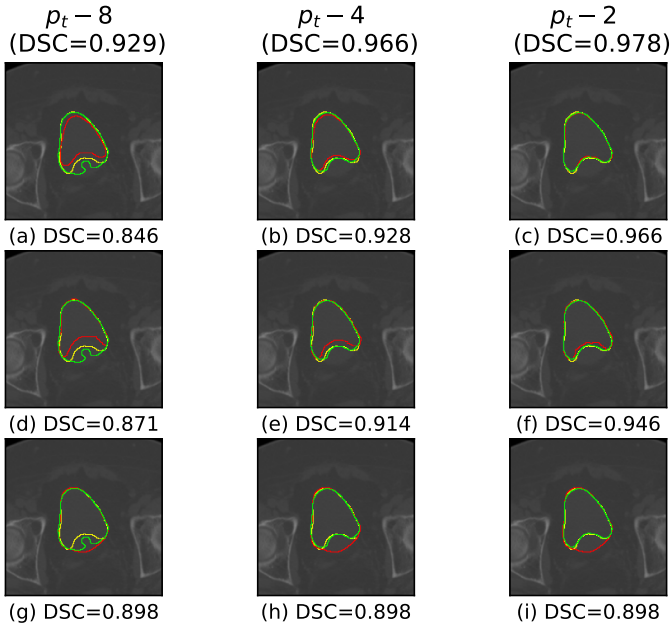

Fig. 4. Ground truth and predicted contours for a given patient at different inter-slices distances. This figure compares the CNN with prior (in green) with the ground truth (in yellow) and with the baselines (in red, each line showing a different baseline algorithm). In the first row, the baseline is the "copy" algorithm, with the DSC between the ground truth and this baseline reported in (a), (b) and (c). In the second line, the baseline is the registration algorithm. In the third row, the baseline is the CNN without prior. The DSC between brackets are computed between the ground truth and the CNN with prior. The first, second and third columns respectively correspond to an inter-slices distance of 8, 4 and 2 mm. DSC: Dice similarity coefficient.

5 Conclusion

We propose a deep CNN network able to segment 2D CT image slices of the bladder, using the manual delineation on an adjacent slice as prior knowledge. This is done by training the u-net network with two input channels. The first channel is used to enter the target image and the second one is used to enter the manual delineation of the adjacent slice given as a binary mask.

The network is trained and tested for an increasing distance between the target and adjacent slices (i.e. from 1 to 10 mm). The Dice similarity coefficient between the thresholded network prediction and the ground truth exceeds 0.9 for all the considered distances and increases as the distance between the slices decreases.

For every distance, the output segmentation outperforms in average the segmentations obtained by a registration-based approach (efficient for small deformations between both slices) and by the u-net network trained and tested without prior knowledge (efficient for large deformations between slices).

In the future, we plan to use the proposed approach on other organs and other applicative contexts where a spatial or temporal contour propagation is needed. We also plan to investigate how the annotation of a few slices within a 3D volume can improve the segmentation performance of a 3D CNN such as 3D u-net [16]. Our code will also be made available at the time of the conference.

Acknowledgments Jean Léger is a Research Fellow of the Fonds de la Recherche Scientifique - FNRS, Eliott Brion's work was supported by FEDER-RW project UserMEDIA, Umair Javaid is a Research Fellow funded by the FNRS Televie grant no. 7.4625.16, John A. Lee and Christophe De Vleeschouwer are Senior Research Associates with the Belgian F.R.S.-FNRS. We thank CHU-UCL-Namur (Dr J.-F. Daisne) as well as CHU-Charleroi (Dr N. Meert) for providing the data.

References

1. Mazurowski, M.A., Buda, M., Saha, A., Bashir, M.R.: Deep learning in radiology: an overview of the concepts and a survey of the state of the art. arXiv preprint arXiv:1802.08717 (2018)
2. Sharp, G., Fritscher, K.D., Pekar, V., Peroni, M., Shusharina, N., Veeraraghavan, H., Yang, J.: Vision 20/20: perspectives on automated image segmentation for radiotherapy. Medical physics **41**(5) (2014)
3. Cha, K.H., Hadjiiski, L., Chan, H.P., Weizer, A.Z., Alva, A., Cohan, R.H., Caoili, E.M., Paramagul, C., Samala, R.K.: Bladder cancer treatment response assessment in CT using radiomics with deep-learning. Scientific reports **7**(1), 8738 (2017)
4. Iglesias, J.E., Sabuncu, M.R.: Multi-atlas segmentation of biomedical images: a survey. Medical image analysis **24**(1), 205–219 (2015)
5. Cremers, D., Rousson, M., Deriche, R.: A review of statistical approaches to level set segmentation: integrating color, texture, motion and shape. International journal of computer vision **72**(2), 195–215 (2007)
6. Heimann, T., Meinzer, H.P.: Statistical shape models for 3D medical image segmentation: a review. Medical image analysis **13**(4), 543–563 (2009)
7. Polan, D.F., Brady, S.L., Kaufman, R.A.: Tissue segmentation of computed tomography images using a random forest algorithm: a feasibility study. Physics in Medicine & Biology **61**(17), 6553 (2016)
8. Luo, S., Hu, Q., He, X., Li, J., Jin, J.S., Park, M.: Automatic liver parenchyma segmentation from abdominal CT images using support vector machines. In: Complex Medical Engineering, 2009. CME. ICME International Conference on. pp. 1–5. IEEE (2009)
9. Hu, Y.C.J., Grossberg, M.D., Mageras, G.S.: Semi-automatic medical image segmentation with adaptive local statistics in conditional random fields framework. In: Engineering in Medicine and Biology Society, 2008. EMBS 2008. 30th Annual International Conference of the IEEE. pp. 3099–3102. IEEE (2008)

10. Tong, T., Wolz, R., Wang, Z., Gao, Q., Misawa, K., Fujiwara, M., Mori, K., Hajnal, J.V., Rueckert, D.: Discriminative dictionary learning for abdominal multi-organ segmentation. Medical image analysis **23**(1), 92–104 (2015)
11. Gao, Y., Shao, Y., Lian, J., Wang, A.Z., Chen, R.C., Shen, D.: Accurate segmentation of CT male pelvic organs via regression-based deformable models and multi-task random forests. IEEE transactions on medical imaging **35**(6), 1532–1543 (2016)
12. Oda, M., Shimizu, N., Karasawa, K., Nimura, Y., Kitasaka, T., Misawa, K., Fujiwara, M., Rueckert, D., Mori, K.: Regression forest-based atlas localization and direction specific atlas generation for pancreas segmentation. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. pp. 556–563. Springer (2016)
13. Litjens, G., Kooi, T., Bejnordi, B.E., Setio, A.A.A., Ciompi, F., Ghafoorian, M., van der Laak, J.A., van Ginneken, B., Sánchez, C.I.: A survey on deep learning in medical image analysis. Medical image analysis **42**, 60–88 (2017)
14. Long, J., Shelhamer, E., Darrell, T.: Fully convolutional networks for semantic segmentation. In: Proceedings of the IEEE conference on computer vision and pattern recognition. pp. 3431–3440 (2015)
15. Ronneberger, O., Fischer, P., Brox, T.: U-net: Convolutional networks for biomedical image segmentation. In: International Conference on Medical image computing and computer-assisted intervention. pp. 234–241. Springer (2015)
16. Çiçek, Ö., Abdulkadir, A., Lienkamp, S.S., Brox, T., Ronneberger, O.: 3D U-Net: learning dense volumetric segmentation from sparse annotation. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. pp. 424–432. Springer (2016)
17. Milletari, F., Navab, N., Ahmadi, S.A.: V-net: Fully convolutional neural networks for volumetric medical image segmentation. In: 3D Vision (3DV), 2016 Fourth International Conference on. pp. 565–571. IEEE (2016)
18. Ibragimov, B., Xing, L.: Segmentation of organs-at-risks in head and neck CT images using convolutional neural networks. Medical physics **44**(2), 547–557 (2017)
19. Kazemifar, S., Balagopal, A., Nguyen, D., McGuire, S., Hannan, R., Jiang, S., Owrangi, A.: Segmentation of the prostate and organs at risk in male pelvic CT images using deep learning. arXiv preprint arXiv:1802.09587 (2018)
20. Roth, H.R., Oda, H., Hayashi, Y., Oda, M., Shimizu, N., Fujiwara, M., Misawa, K., Mori, K.: Hierarchical 3D fully convolutional networks for multi-organ segmentation. arXiv preprint arXiv:1704.06382 (2017)
21. Larsson, M., Zhang, Y., Kahl, F.: Robust abdominal organ segmentation using regional convolutional neural networks. In: Scandinavian Conference on Image Analysis. pp. 41–52. Springer (2017)
22. Milletari, F., Rothberg, A., Jia, J., Sofka, M.: Integrating statistical prior knowledge into convolutional neural networks. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. pp. 161–168. Springer (2017)
23. Trullo, R., Petitjean, C., Ruan, S., Dubray, B., Nie, D., Shen, D.: Segmentation of organs at risk in thoracic CT images using a sharpmask architecture and conditional random fields. In: Biomedical Imaging (ISBI 2017), 2017 IEEE 14th International Symposium on. pp. 1003–1006. IEEE (2017)
24. Klein, S., Staring, M.: elastix, the manual, 2018. Available at http://elastix.isi.uu.nl/download/elastix-4.9.0-manual.pdf