Automatic Segmentation of 3D MRI for Cervical Cancer Radiation Therapy with Transfer Learning

Sze-Nung Leung[1] SZENUNG.LEUNG@UQ.EDU.AU
Wei Dai[1] WEI.DAI@UQ.EDU.AU
[1] *University of Queensland, Australia*
Jason A. Dowling[2] JASON.DOWLING@CSIRO.AU
[2] *CSIRO Australian e-Health Research Centre, Australia*
Karen Lim[4] KAREN.LIM@HEALTH.NSW.GOV.AU
Tony Young[3,4,5] TONY.YOUNG@HEALTH.NSW.GOV.AU
[3] *Institute of Medical Physics, University of Sydney, Australia*
Lois Holloway[4,5,6] LOIS.HOLLOWAY@HEALTH.NSW.GOV.AU
[4] *Cancer Therapy Centre, Liverpool Hospital, Australia*
[5] *Liverpool and Macarthur Cancer Therapy Centres and Ingham Institute, Australia*
[6] *South Western Clinical School, University of New South Wales, Australia*
Shekhar S. Chandra[1] SHEKHAR.CHANDRA@UQ.EDU.AU

Editors: Under Review for MIDL 2022

Abstract

Cervical cancer is the fourth most common cancer in women worldwide. Treatment of the disease includes radiation therapy treatment, requiring precise organ delineation for treatment planning. Improvement of the segmentation process for the tumour site and nearby organs could limit treatment side effects and improve treatment planning efficiency. Compared with segmentation for other anatomical locations, Magnetic Resonance Imaging (MRI) for cervical cancer is challenging based on the limited amount of training data and large inter-patient shape variation for organs close to the tumour (organs-at-risk). The proposed approach included transfer learning with male pelvis data with similar field of view (FOV), with the addition of dilated layers within the 3D U-Net structure to improve segmentation performances. The approaches were compared to the 3D U-Net which was widely used in recent studies in image segmentation for cervical cancer MRI or computed tomography (CT). The data used consisted of 42 images obtained from 20 patients with stage IB to IVB cervical cancer across a maximum of 7 weeks of radiation therapy with manually contoured labels including the bladder, cervix, gross tumour volume (GTV), uterus and rectum. The studied approaches accounted for the small dataset and large range of tumour size. Outcomes were evaluated based on the Dice Similarity Coefficients (DSC), the Hausdorff Distance (HD) and the Mean Surface Distance (MSD). The approaches evaluated were shown to have improved segmentation outcomes and required a significantly lower number of model parameters, reducing computer cost and memory.

Keywords: 3D multi-organ segmentation, MRI, Cervical Cancer, CNN, transfer learning

1. Introduction

Cervical cancer is a common disease causing over 300,000 deaths worldwide each year. Computed Tomography (CT) scans and Magnetic Resonance Imaging (MRI) are used for

radiotherapy treatment planning, and accurate target and organ-at-risk delineation is essential. Cervical cancer can extend to the pelvic side wall or invade adjacent organs such as the bladder or rectum. CT scans are useful for staging and follow-up in patients, however, CT provides poor soft-tissue contrast, limiting its use for local staging. MRI provides improved soft tissue contrast and is useful for defining the extent of tumour involvement in terms of contour and volume (Veera et al., 2019). External beam radiation therapy planning contains some uncertainty caused by organ movements and volume changes, in addition to patient geometry differences. A margin for the planning dose target volume is generally expanded around the tumour area to ensure it receives adequate radiation dose, resulting in dose being delivered to normal tissues as a side effect (Jadon et al., 2014).

Precise delineation of organ boundaries from medical imaging is essential for accurate treatment application, reducing side effects. Conventionally, manual segmentation is used, which has limitations including inconsistency caused by inter-observer differences, and time inefficiency (the contouring for a single cervical cancer patient takes 90 to 120 minutes to be completed (Liu et al., 2020)). The process also require specialised staff with a high level of expertise. For cervical cancer, image based automatic segmentation is particularly challenging due to the large inter-patient shape variation and the limited data sources. In this paper, convolutional neural network (CNN) based 3D networks were investigated for the application of cervical cancer segmentation. Contribution of this paper included:

1. A new clinical 3D MRI dataset for the female pelvis organs from the study in Veera et al. (2019) was investigated for full 3D automatic segmentation.

2. A modified version of CAN3D (Dai et al., 2021) was introduced. The proposed model achieved same level or better results with respect to the standard 3D U-Net with significantly less training time computer power required.

3. Transfer learning from male pelvis region MRI scans with a similar field of view (FOV) from a previous prostate cancer study (Dowling and Greer, 2021) was applied in this paper. It improved model performance, suggesting an alternative approach for increasing training data for cervical cancer MRI segmentation with small dataset in addition to data augmentation.

2. Related work

2.1. Cervical Cancer MRI

There are few studies which have been explored on pelvic organ segmentation for cervical cancer treatment planning.

One of these studies reported the development of a registration framework for the application of bladder and clinical target volume (CTV) segmentation in pelvic MRI. A statistical shape model was used as a constraint during optimization (Berendsen et al., 2013) with a resulting Dice Similarity Coefficient (DSC) of 0.67/0.73 and 0.55/0.57 (mean/ median) for the bladder and CTV respectively. Considering the large shape variation of the cervix-uterus shape and position, another study proposed a patient specific predictive model for it based on the relationship between bladder filling and the motion of cervix-uterus using deformation registration (Bondar et al., 2011). Individual CT scans was needed for each

patient (with full and empty bladder) to generate the model and its approach significantly reduced the margin required to account for motion uncertainties. These approaches all required prior development of a constraint model which were not an efficient solution with the low accessibility of cervical cancer MRI data.

Deep learning approach with a CNN has been widely explored in recent studies. The use of a 2D and 3D U-Net was tested with segmentation for the gross tumour volume (GTV) from 98 cervical cancer MRI images with stage IB to IVA. The resulting DSC ranged from 0.13 to 0.93 with mean and median of 0.77 and 0.83 (Kano et al., 2021). A U-Net was also tested for uterine segmentation with or without the presence of major uterine disorders. The results were satisfying despite the presence of uterine disorders (Kurata et al., 2019). In a study by Rigaud et al. (2021), the 2D DeepLabV3+ and 3D 2-steps U-Net in RayStation 9B with RayLearner was used for the segmentation of 12 anatomic structure from 408 cervical cancer CT scans where the segmentation results for the cervix-uterus, parametrium, and 2-cm upper part of the vagina was used to define the primary CTV. The 2D and 3D model had a mean DSC of 0.81 and 0.82 respectively regardless of inter-observer variability. Another study proposed a segmentation model for organs including the bladder, bone marrow, femoral heads, rectum, small intestine and spinal cord from 105 CT images (Liu et al., 2020). The structure of the network was based on 2D U-Net, segmenting CT images slice by slice. Apart from the typical U-Net, the convolutional layers were replaced by context aggregation block with convolution followed by Squeeze and Excitation to re-weight each organ mask (Liu et al., 2020). The context aggregation block used dilated convolutions with varying dilation rate to collect the features in a larger receptive field, along with normal convolution of different kernel size to reduce "gridding" artefacts of dilated convolutions. The proposed network had an improved result compared with the application of U-Net (Ronneberger et al., 2015).

For organ segmentation in cervical cancer MRI, U-Net based structured network has been used for both 2D and 3D applications. Due to the large shape variability of the cervix and GTV, most studies have focused on nearby organs that have a more predictable structure including the bladder, rectum and bone structures. 2D models were used in most studies (with 3D data requiring high computer cost and memory).

2.2. Medical Image Segmentation

For medical image segmentation, different network structures were explored. A 3D U-Net was developed for the application of 3D volumetric segmentation (Cicek et al., 2016). A modified 3D U-Net was developed in Isensee et al. (2018) with residual connections. The proposed network included deep supervision in the localization pathway and achieved promising results on the BraTS 2017 challenge. Dilated convolution was widely used to collect contextual information without losing resolution (Hamaguchi et al., 2018; Li et al., 2018). In the study of (Yu and Koltun, 2016), dilated convolution was used for context aggregation and was proposed for the purpose of semantic segmentation. Semantic segmentation generally involving the reduction of resolution to collect image information. The study proposed an approach that could extract contextual information in combination with the full-resolution output. Dilated convolution supports exponential expansion of the receptive field and eliminated the need of compressing the input (Yu and Koltun, 2016). In

the study by (Dai et al., 2021), CAN3D was developed with the use of dilated convolution layers (DCL) for 3D volumes, reserving image information. The network was applied for prostate cancer MRI data and had shown improvement in results and efficiency compared to state-of-art approaches including U-Net.

In this paper, 3D U-Net and a modified CAN3D were investigated with transfer learning and applied to cervical cancer MRI data. The aim was to account for the limited source of MRI data in this domain and to allow direct application on 3D images with reasonable computational cost.

3. Material and Method

3.1. Data

Data used in this study was obtained from 22 adult patients with stage IB-IVB cervical cancer with a macroscopic tumour. They were recruited from Liverpool and Macarthur Cancer Therapy Centre in Sydney, Australia. Ethics approval and informed consent was obtained from all patients. MRI scans were collected during primary radiation therapy with a maximum of 7 scans per patients. Two patient results were of poor quality with low visibility of the target volumes and were removed from this study, resulting in a total of 42 labelled datasets. MRI scans were acquired on a 70cm bore MAGNETOM Skyra 3T scanner. The MR sequence for each scan was a 2D axial T2-weighted turbo-spin echo with 2mm slices and a single concatenation (Veera et al., 2019). This dataset was obtained in 2013 to 2019 for the purpose of a clinical study and has been utilised in this work for the first time for automatic segmentation. The dataset was selected to investigate the application of deep learning techniques on a small medical imaging dataset with large structural variations.

3.2. Network Structure

Different image segmentation approaches were applied on the cervical cancer MRI dataset. Semantic segmentation of the target structures (bladder, cervix, GTV, uterus and rectum) was performed from the 3D images and evaluated.

3.2.1. 3D U-Net

A U-Net consists of contraction layers to condense and extract pattern and information from the input, and a decoder as the reconstruction process to obtain the final result (Ronneberger et al., 2015). Skip connections are present to reserve information lost in the encoder part and to be used for the reconstruction process. The 3D U-Net structure introduced in Cicek et al. (2016) was applied in this paper with filter numbers reduced by a factor of 2 due to computer memory limitations.

3.2.2. CAN3D+

A mixed version of CAN3D (Dai et al., 2021) (Figure 3) and 3D U-Net was also investigated. The architecture of the network is provided in Figure 1. The standard convolution block includes a 3D convolution layer, instance normalisation layer, and a ReLU activation layer. A convolution layer with strides of 2 was used for down-sizing instead of a max-pooling

layer to reduce computational cost during training with the large 3D data size (Milletari et al., 2016).

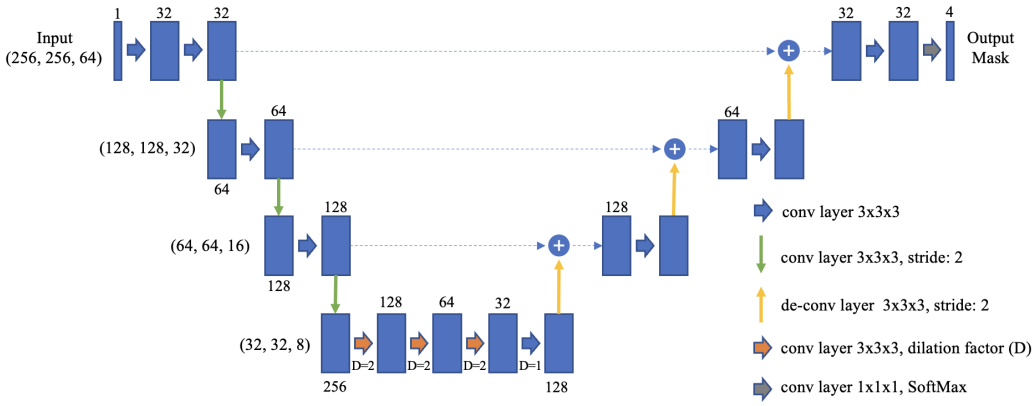


Figure 1: Adjusted CAN3D architecture. Standard convolution block include a convolution layer, instance normalisation and ReLU.

In contrast to 3D U-Net, the last down-sampling level was removed. Three dilated convolution layers (DCL) were present between the encoder and decoder. The application of dilated convolution was introduced to increase receptive field of the convolution operation for aggregating contextual information without resolution reduction (Yu and Koltun, 2016; Li et al., 2018; Dai et al., 2021). Despite its advantages, it was proven to cause "gridding artefacts" as network outputs were derived from separated input voxels (Yu et al., 2017; Wang and Ji, 2018). In this model, DCLs were of the same dilation rates instead of the progressive DCLs used in CAN3D. The decreasing filter numbers give the latter layers less importance. With the layers applied consecutively, the receptive field was still increased while limiting the gap between voxels in which the output convolution results were obtained. The DCLs were followed by one normal convolution layer and 3 reconstruction layers. Reconstruction layers consists of one up-convolution layer with strides of 2, followed by 2 convolution layers. The up-sampled input of the layer was joined by the output of the contraction layer with corresponding sizes to regain lost information during the encoder part. A last convolution layer with kernel size of 1 and channel size of the number of organ-at-risk for segmentation was included to obtain the final result. The last layer had an activation function of softmax, which output the probability of each voxel to belong to each output channel.

3.3. Transfer Learning

Cervical cancer MRI were hard to obtain with the number of organs involved and resources required. In this paper, data from 20 patients was available for the development of an automated segmentation network. This dataset has a wide range of bladder, tumour and rectum sizes with different patient conditions and stage of cancer at the time of scan acquisition. With a similar acquisition protocol, prostate MRI data has been widely researched and the

data is relatively more accessible. A set of prostate cancer MRI with a similar FOV was used to train the model to learn the basic structure of the region before the model was trained with the cervical cancer MRI data (Appendix B).

Prostate MRI data used for transfer learning consists of 211 3D images from 38 patients diagnosed with localised prostate cancer (Dowling and Greer, 2021). The original data size was 256x256x128 which was cropped to 256x256x64 to fit to our model. Manual contours were provided for the bladder, body, bone, rectum and prostate. This dataset has used in previous studies by Dai et al. (2021).

3.4. Implementation

The data were divided into 5 sets with 7-9 images using a 5-fold cross validation. Each of the 5 sets of data was used as the test set once (Appendix A Figure 4). 10% of the data was used as the validation data and model parameters were chosen based on its validation loss. The average and median of the evaluation metric results were obtained to determine the performance of our model. The training set went through data augmentation 11 times with random operations including elastic deformation, affine wrapping, uni-axial rotation and shift, resulting in 420-444 set of 3D images used for model training. The dataset was provided with manually contoured labels of the bladder, cervix, GTV, uterus and rectum. Patient data included GTV structures covering different percentage of the cervix and/or the uterus. To account for all cases in the study, the cervix, GTV and uterus structure were merged for network training purpose. The combined structure assists radiation oncologists in the identification of a region within MRI which will include the cancerous tumour. The inclusion of organs-at-risk other than the tumour volume are important for generating the treatment dose plan or to identify significant changes in anatomy where dose will need to be re-planned. N4 bias field correction (Tustison et al., 2010) was performed for pre-processing of data. Input images were all cropped and resized to the image size of 256x256x64 with a voxel size of 1.64x1.64x3mm.

For network training, the Dice loss function was used for our model. An alternative representation of the loss function was shown as Equation (1) with p as ground truth and q as predicted results (Cha, 2007). Adam optimization algorithm with a learning rate of 0.001 was applied to the model with a batch size of 1 to limit computational cost while optimising model features. The model was implemented on TensorFlow 2.7.0 and the training processes were conducted on a NVIDIA SXM2 Tesla 32 GB V100 GPU.

$$D_{loss} = \frac{\sum (p-q)^2}{\sum p^2 + \sum q^2}. \tag{1}$$

3D U-Net and CAN3D+ were both investigated on this cervical cancer patient dataset. The Dice Similarity Coefficient (DSC), Hausdorff Distance (HD) and Mean Surface Distance (MSD) were used as evaluation metrics for the models against the manual contours provided. As different organ structures of the model have varying volumes, the HD and MSD were also used along with DSC for further evaluation (Prados et al., 2017).

4. Results

To demonstrate improvement due to transfer learning (TL) with the prostate data, the standard U-Net and the CAN3D+ was trained from scratch, and compared with the same network first trained with the prostate data for 10 epochs. The trained model weights were transferred into a new model for the cervix dataset. DSC results comparison was displayed in Figure 2. Detailed results including the average and median value of DSC, HD and MSD were displayed in Appendix B Table 1.

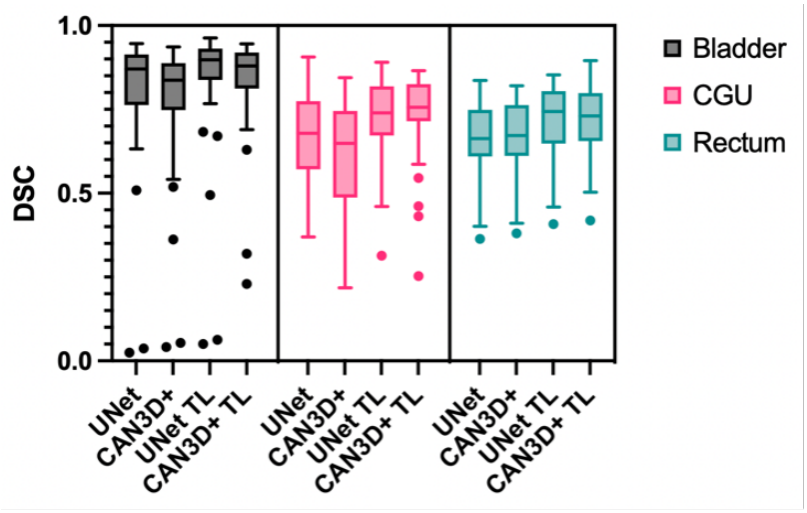


Figure 2: Boxplot of the DSC results for the bladder, combined cervix-GTV-uterus (CGU) and rectum from the proposed methods.

All networks with and without TL were trained with the same cervix data, with the networks with TL showing an improvement in model performance. The prostate dataset included contours of the body, bone, bladder, rectum and prostate. It provided extra training data for the model to learn the overall structure of the pelvic region. The model weights trained from the prostate data were transferred into 5 new U-Net models from the 5-fold validation for cervix data training. This allowed a more efficient training process of the model while preventing over-fitting of the available cervix data.

From Appendix B Table 1, the quantitative results of the CAN3D+ are also displayed. The processing time and the number of model parameters for the models were 1.83s/step and 22,713,508, and 1.01s/step and 4,899,942 for U-Net and CAN3D+ respectively. 55% of the training time and 0.2157 of computer memory and power was required to train the CAN3D+ compared to U-Net. CAN3D+ without TL demonstrated worse performance for the bladder and CGU structure with the limited dataset, but had better results for the prostate dataset (Appendix B Table 2), and the best overall results when TL was applied. The CAN3D+ with TL demonstrated an improved output for the GTV structure from the standard U-Net with similar result for the bladder and rectum. While the U-Net with TL had a slightly better result than CAN3D+ with TL for the rectum, CAN3D+ results had

a better distribution and was able to achieve a similar level of output as the U-Net with significantly less resources.

Inter-observer contouring variability of this dataset was studied in Veera et al. (2019). The mean DSC and average mean absolute surface distance MASD (mm) was 0.48/4.59, 0.81/3.11 and 0.82/2.23 for the cervix, GTV and uterus respectively. The result of the CGU structure from the proposed model was within the range of the inter-observer variability.

Within the dataset, GTV ranged from 11.42cm^3 to 712.31cm^3 (median 111.7cm^3), which caused some images to have poor prediction results and resulted in increased standard deviation. The same situation applied to bladder and rectum where bladder volume ranged from 30.08cm^3 to 721.8cm^3 (median 200cm^3) and 19.3cm^3 to 213.45cm^3 (median 57.95cm^3) for rectum. The segmentation results were improved with an inflated bladder and rectum, and in cases where the gross tumour was close to the median size. Output examples are shown in Appendix B Figure 5. These examples were chosen to show results with full or empty bladder and rectum, along with different gross tumour volumes. A good (top row) and poor (bottom row) result was chosen as example to show the differences between the approaches. CAN3D+ TL results were smoother and more concise compared to the other approaches.

Overall results of the CAN3D+ have demonstrated a significant improvement in the efficiency of the model while producing better or similar results compared to the U-Net model with much less resources. However, model performance were limited by the shape variability of the targeted structures. Shape variability was present between data due to inter-patient differences, range of cancer stages, and patients' condition at the time of data retrieval. Future analysis on this issue is likely to help further improve model performances.

5. Conclusion

Several methods were investigated for 3D segmentation of cervical cancer MRI. In addition to augmented images, transfer learning provided extra training data without the risk of over-fitting. It was shown to be able to account for the limited dataset size that is typical for clinical studies and which is generally an issue for cervical cancer image processing studies. Compared to the standard 3D U-Net, the modified CAN3D required significantly less training time and computer cost while achieving similar or better results. The results were largely affected by the shape variability of the targeted structures and future analysis focused on this issue would be beneficial.

References

Floris F. Berendsen, Uulke A. van der Heide, Thomas R. Langerak, and Alexis N.T.J. Kotte. Free-form image registration regularized by a statistical shape model: application to organ segmentation in cervical mr. *Computer Vision and Image Understanding*, 117: 1119–1127, 2013.

Luiza Bondar, Mischa Hoogeman, Jan Willem Mens, Glenn Dhawtal, Ilse de Pree, Rozilawati Ahmad, Sandra Quint, and Ben Heijmen. Toward an individualized target motion management for imrt of cervical cancer based on model-predicted cervix–uterus shape and position. *Radiotherapy and Oncology*, 99:240–245, 2011.

Sung-Hyuk Cha. Comprehensive survey on distance/similarity measures between probability density functions. *International Journal of Mathematical Models and Methods in Applied Sciences*, 1, 2007.

Ozgun Cicek, Ahmed Abdulkadir, Soeren S. Lienkamp, Thomas Brox, and Olaf Ronneberger. 3d u-net: Learning dense volumetric segmentation from sparse annotation. 2016.

Wei Dai, Boyeong Woo, Siyu Liu, Matthew Marques, Craig B. Engstrom, Peter B. Greer, Stuart Crozier, Jason A. Dowling, and Shekhar S. Chandra. CAN3D: Fast 3D Medical Image Segmentation Via Compact Context Aggregation. *IEEE International Symposium on Biomedical Imaging*, 2021. doi: https://doi.org/10.1109/ISBI48211.2021.9433784.

Jason Dowling and Peter Greer. Labelled weekly mr images of the male pelvis. 2021. doi: https://doi.org/10.25919/45t8-p065.

Ryuhei Hamaguchi, Aito Fujita, Keisuke Nemoto, Tomoyuki Imaizumi, and Shuhei Hikosaka. Effective use of dilated convolutions for segmenting small object instances in remote sensing imaging. *IEEE Winter Conference on Applications of Computer Vision*, 2018.

Fabian Isensee, Philipp Kickingereder, Wolfgang Wick, Martin Bendszus, and Klaus H. Maier-Hein. Brain tumor segmentation and radiomics survival prediction: Contribution to the brats 2017 challenge. 2018.

R. Jadon, C.A. Pembroke, C.L. Hanna, N. Palaniappan, M. Evans, A.E. Cleves y, and J. Staffurth. A systematic review of organ motion and image-guided strategies in external beam radiotherapy for cervical cancer. *Clinical Oncology*, 26:185–196, 2014.

Yosuke Kano, Hitoshi Ikushima, Motoharu Sasaki, and Akihiro Haga. Automatic contour segmentation of cervical cancer using artificial intelligence. *Journal of Radiation Research*, 62(5):934–944, 2021.

Yasuhisa Kurata, Mizuho Nishio, Aki Kido, Koji Fujimoto, Masahiro Yakima, Hiroyoshi Isoda, and Kaori Togashi. Automatic segmentation of the uterus on mri using a convolutional neural network. *Computers in Biology and Medicine*, 114, 2019.

Xia Li, Jianlong Wu, Zhouchen Lin, Hong Liu, and Hongbin Zha. Recurrent squeeze-and-excitation context aggregation net for single image deraining. 2018.

Zhikai Liu, Xia Liu, Bin Xiao, Shaobin Wang, Zheng Miao, Yuliang Sun, and Fuquan Zhang. Segmentation of organs-at-risk in cervical cancer ct images with a convolutional neural network. *Physica Medica*, 69:184–191, 2020.

Fausto Milletari, Nassir Navab, and Seyed-Ahmad Ahmadi. V-net: Fully convolutional neural networks for volumetric medical image segmentation. 2016.

Ferran Prados, John Ashburner, Claudia Blaiotta, Tom Brosch, Julio Carballido-Gamio, Manuel Jorge Cardoso, Benjamin Conrad, Esha Datta, Gergely David, Benjamin

De Leener, Sara Dupont, Patrick Freund, Claudia Gandini Wheeler-Kingshott, Francesco Grussu, Roland Henry, Bennett Landman, Emil Ljungberg, Bailey Lyttle, Sebastien Ourselin, and Julien Cohen-Adad. Spinal cord grey matter segmentation challenge. *NeuroImage*, 152, 03 2017. doi: 10.1016/j.neuroimage.2017.03.010.

Bastien Rigaud, Brian M. Anderson, Zhiqian H. Yu, Maxime Gobeli, Guillaume Cazoulat, Jonas Soderberg, Elin Samuelsson, David Lidberg, Christopher Ward, Nicolette Taku, Carlos Cardenas, Dong Joo Rhee, Aradhana M. Venkatesan, Christine B. Peterson, Laurence Court, Stina Svensson, Fredrik Lofman, Ann H. Klopp, and Kristy K. Brock. Automatic segmentation using deep learning to enable online dose optimization during adaptive radiation therapy of cervical cancer. *International Journal of Radiation Oncology*, 109(4), 2021.

Olaf Ronneberger, Philipp Fischer, and Thomas Brox. U-net: Convolutional networks for biomedical image segmentation. pages 234–241, 2015.

Nicholas J. Tustison, Brian B. Avants, Philip A. Cook, Yuanjie Zheng, Alexander Egan, Paul A. Yushkevich, and James C. Gee. N4itk: Improved n3 bias correction. *IEEE Trans Med Imaging*, 29(6):1310–1320, 2010.

Jacqueline Veera, Karen Lim, Jason A Dowling, Chelsie O'Connor, Lois C Holloway, and Shalini K Vinod. Dedicated mri simulation for cervical cancer radiation treatment planning: Assessing the impact on clinical target volume delineation. *J Med Imaging Radiat Oncol*, 63(2):236–243, Apr 2019.

Zhengyang Wang and Shuiwang Ji. Smoothed dilated convolutions for improved dense prediction. *Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, 62:4140–4159, Jul 2018.

Fisher Yu and Vladlan Koltun. Multi-scale context aggregation by dilated convolutions. 2016.

Fisher Yu, Vladlen Koltun, and Thomas Funkhouser. Dilated residual networks. 2017.

Appendix A. Experiment Implementation

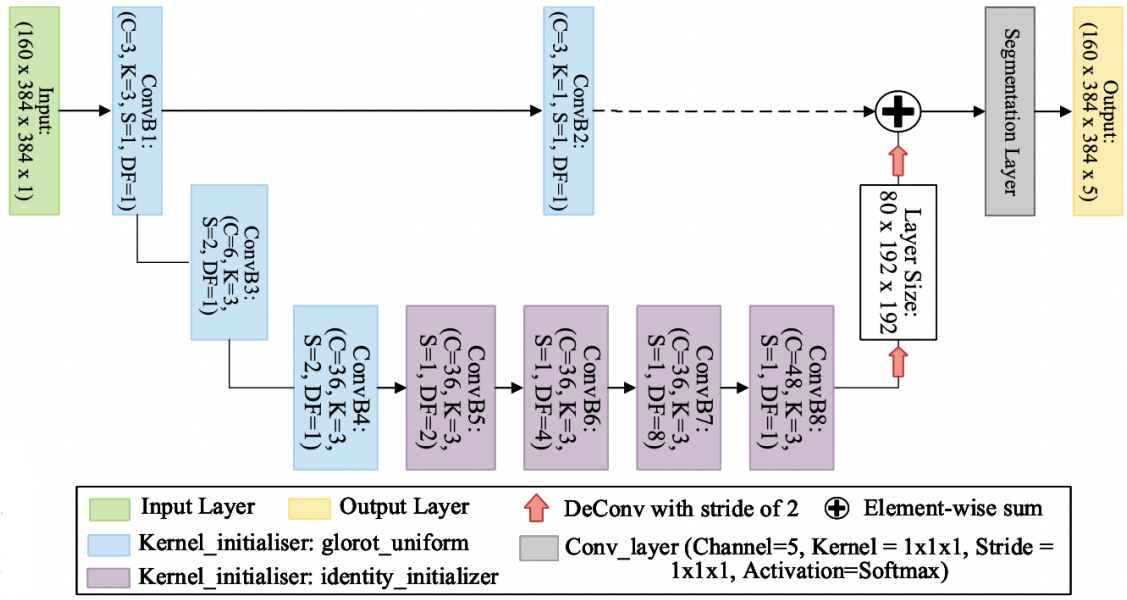


Figure 3: CAN3D architecture. (Dai et al., 2021)

Patient	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
Set 1																				
Set 2																				
Set 3																				
Set 4																				
Set 5																				

Figure 4: Data split for 5-fold cross-validation. Orange: testing/ Blue: validation/ Blank: training

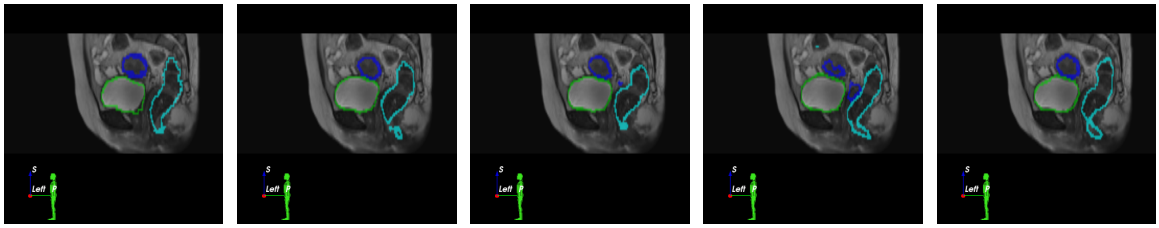
Appendix B. Additional Results

Table 1: Evaluation of the proposed methods. The models were compared with average and median of DSC, HD and MSD (mm) for the bladder, combined cervix-GTV-uterus (CGU) and rectum ± standard deviation. The best values are in bold.

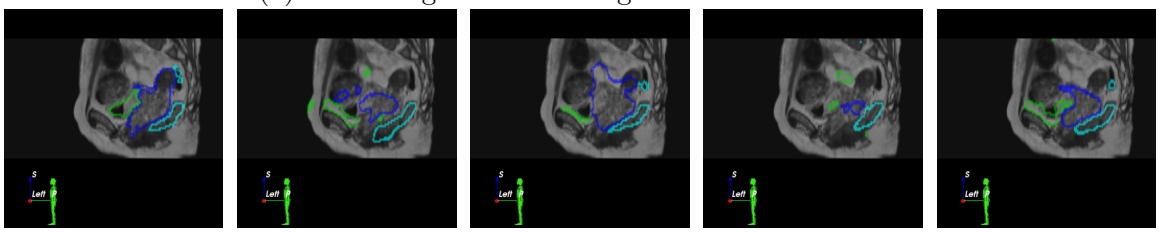
Structure	**Metric**	**U-Net**	**U-Net TL**	**CAN3D+**	**CAN3D+ TL**
Bladder	DSC	$0.794/0.870_{\pm 0.210}$	$\mathbf{0.829/0.898}_{\pm 0.206}$	$0.761/0.837_{\pm 0.217}$	$\mathbf{0.830/0.880}_{\pm 0.154}$
	HD	$32.65/21.40_{\pm 36.25}$	$\mathbf{18.63/9.24}_{\pm 22.51}$	$34.89/26.04_{\pm 33.26}$	$26.97/10.66_{\pm 36.78}$
	MSD	$2.91/1.47_{\pm 2.84}$	$\mathbf{1.71/1.10}_{\pm 1.41}$	$3.51/1.96_{\pm 3.71}$	$3.27/1.30_{\pm 5.45}$
CGU	DSC	$0.676/0.678_{\pm 0.128}$	$0.727/0.740_{\pm 0.126}$	$0.610/0.649_{\pm 0.168}$	$\mathbf{0.734/0.756}_{\pm 0.132}$
	HD	$\mathbf{31.36}/29.59_{\pm 15.12}$	$39.35/35.96_{\pm 28.14}$	$33.04/\mathbf{28.40}_{\pm 23.02}$	$39.80/31.14_{\pm 35.52}$
	MSD	$\mathbf{3.07}/2.59_{\pm 1.65}$	$3.41/2.67_{\pm 2.45}$	$3.60/2.63_{\pm 2.97}$	$3.78/\mathbf{2.50}_{\pm 3.46}$
Rectum	DSC	$0.661/0.663_{\pm 0.113}$	$\mathbf{0.716/0.744}_{\pm 0.116}$	$0.672/0.672_{\pm 0.101}$	$0.714/0.731_{\pm 0.109}$
	HD	$45.93/46.40_{\pm 27.82}$	$\mathbf{28.23/18.90}_{\pm 30.37}$	$45.23/26.84_{\pm 39.15}$	$31.77/19.61_{\pm 38.04}$
	MSD	$3.02/2.27_{\pm 1.62}$	$\mathbf{2.25/1.62}_{\pm 1.53}$	$2.77/2.16_{\pm 1.90}$	$2.31/1.75_{\pm 1.54}$
Training speed		1.83s/step		1.01s/step	
Model parameters		22,713,508		4,899,942	

Table 2: Evaluation of the investigated model trained with prostate dataset. The models were compared with average and median DSC. The best values are in bold.

Structure	**U-Net**	**CAN3D+**
Body	$0.979/0.979_{\pm 0.005}$	$\mathbf{0.983/0.983}_{\pm 0.004}$
Bone	$0.867/0.873_{\pm 0.029}$	$\mathbf{0.905/0.908}_{\pm 0.020}$
Bladder	$0.914/0.936_{\pm 0.076}$	$\mathbf{0.928/0.950}_{\pm 0.066}$
Rectum	$0.839/0.860_{\pm 0.068}$	$\mathbf{0.881/0.899}_{\pm 0.055}$
Prostate	$0.824/0.842_{\pm 0.090}$	$\mathbf{0.851/0.866}_{\pm 0.050}$



(a) Patient aged 48 with stage IIB cervical cancer.



(b) Patient aged 59 with stage IVA cervical cancer.

Ground Truth U-Net U-Net with TL CAN3D+ CAN3D+ with TL

Figure 5: Example output from the proposed methods on two patients. Light green: bladder/ dark blue: cervix-GTV-uterus/ light blue: rectum. (a) Example with inflated bladder and rectum, and average sized CGU. Overall DSC >0.66. (b) Example with flat bladder and large tumour size. Bladder DSC ranged from 0 to 0.32 and CGU DSC ranged from 0.50 to 0.72.