Journal of Medical Systems manuscript No.
(will be inserted by the editor)

Versatile convolutional networks applied to Computed Tomography and Magnetic Resonance image segmentation

Gonçalo Almeida · João Manuel R.S. Tavares

Received: date / Accepted: date

Abstract Medical image segmentation has seen positive developments in recent years but remains challenging with many practical obstacles to overcome. The applications of this task are wide-ranging in many fields of medicine, and used in several imaging modalities which usually require tailored solutions.

Deep learning models have gained much attention and have been lately recognized as the most successful for automated segmentation. In this work we show the versatility of this technique by means of a single deep learning architecture capable of successfully performing segmentation on two very different types of imaging: computed tomography and magnetic resonance.

The developed model is fully convolutional with an encoder-decoder structure and high-resolution pathways which can process whole three-dimensional volumes at once, and learn directly from the data to find which voxels belong to the regions of interest and localize those against the background.

The model was applied to two publicly available datasets achieving equivalent results for both imaging modalities, as well as performing segmentation of different organs in different anatomic regions with comparable success.

Keywords Computer Vision · Deep Learning · Medical Imaging · Segmentation

Gonçalo Almeida
Instituto de Ciência e Inovação em Engenharia Mecânica e Engenharia Industrial, Faculdade de Engenharia, Universidade do Porto, Rua Dr. Roberto Frias, s/n, 4200-465 PORTO, PORTUGAL
E-mail: galmeida@inegi.up.pt

João Manuel R.S. Tavares
Instituto de Ciência e Inovação em Engenharia Mecânica e Engenharia Industrial, Departamento de Engenharia Mecânica, Faculdade de Engenharia, Universidade do Porto, Rua Dr. Roberto Frias, s/n, 4200-465 PORTO, PORTUGAL
E-mail: tavares@fe.up.pt **(corresponding author)**

1 Introduction

Segmentation of medical images is useful in a myriad of disciplines in medicine, ranging from screening and diagnosis to treatment and follow-up. In this process, one or more regions of interest (ROI) are extracted from the patient's anatomical information, in order to guide medical care. It is applied to practically all types of medical images: various forms of X-ray imaging such as mammograms and chest X-rays, Computed Tomography (CT), Magnetic Resonance Imaging (MRI), ultrasounds, nuclear medicine scans such as scintigraphy and Positron Emission Tomography (PET), and also in natural photographs for surgery planning or forensic medicine. Typically, this is achieved by means of manual delineation by a medical practitioner.

In this work, the focus is on radiographic images of two types in particular: computed tomography and magnetic resonance imaging. Although these two image modalities are used in the same anatomical regions and often for similar purposes, their working principles and methods for acquiring information are very dissimilar, resulting in images that are very different, highlighting different cellular tissues, as evidenced in Figure 1. Usually MRI offers higher resolution and better contrast for soft-tissues, while CT helps identify high density structures easily and is much less expensive. Owing to these differences, CT and MRI can be used complementary for the same patient, highlighting the need to have segmentation methods accurate for both image modalities.

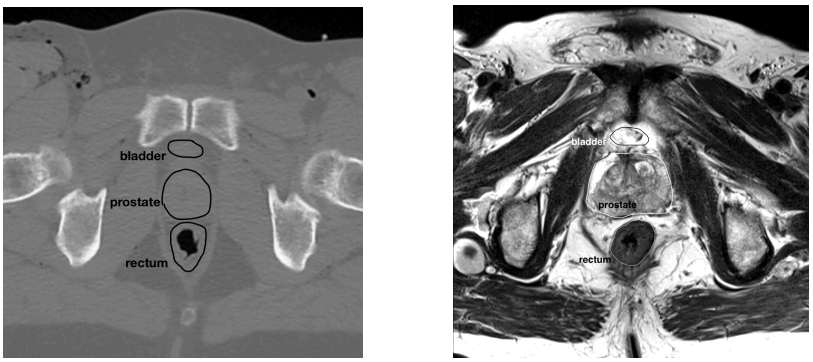

Fig. 1 Examples of CT (on the left) [1] and MR (on the right) [2] images, at roughly the same anatomical position for comparison purposes (these images belong to two different subjects).

The simplest computational segmentation technique consists of a thresholding method, where pixels above a certain value are classified as object of interest and the remaining as belonging to the background. Although simple, this method is often capable of achieving very good results. Over the years, a myriad of rule-based as well as statistical models have been developed that gradually improved automatic and semi-automatic segmentation, tailored to specific uses and specific images. Recently, as in other fields of computational

vision, the interest has shifted to deep learning, mostly due to its performance [3].

The vision for widespread use of automatic segmentation in medicine offers immense gains in productivity and efficiency: no uncertainties associated with manual delineation and practitioner variability, ensuring consistently accurate anatomic encompassing of the target volume, increasing treatment success and patient survival; increase in physician time to spend with patients or studying and researching; less time between first patient encounter and treatment start, which improves outcomes.

This article is structured as follows: after this introduction, there is a section covering related work, with an overview of medical image segmentation methods. Afterwards, the methods section gives details about the developed deep learning model. Section 4 describes the results obtained with our model in the segmentation of two public CT and MRI datasets, and presents comparative experiments with state of the art models. Lastly, in section 5, a discussion is provided and final conclusions are drawn.

2 Related Work

Although automatic medical image segmentation has been studied for many years and significant advances have been accomplished, it remains an active research field with no definite general solution. The main challenges arise from inhomogeneity in images, low contrast, noise, artifacts and human factors, as well as details associated with image acquisition.

Clustering methods have been used for image segmentation as the purpose of aggregating coherent regions of the image is satisfied. However, the precise localization of irregular borders poses problems, together with the need to previously determine number of clusters and centroids [4, 5].

The deformable model offers good accuracy tailored to boundaries, such as the active contour method [6]. Other approaches focus on regions, such as region-growing and region merging, with complex algorithms tailored to remove noise and enhance boundary homogeneity [7].

Machine learning (ML) experts were typically constrained to building models in two steps: feature representation, where a complex, carefully engineered feature extraction process was performed, followed by a predictive model [8]. But the idea of a computational program that could learn representation from dataset to final result had been sought. The breakthrough with deep learning (DL) is that no feature extraction is required. DL models can learn to extract the features most important from the training data itself [9]. These have become the state of the art in medical image analysis, evidenced by most of the challenges being populated with these methods in all the top positions nowadays. However, it is worth understanding that this only became true recently. It was only in 2017 that a 3D fully Convolutional Neural Network (fCNN) architecture by Yu et al.[10] captured first place in the PROMISE12

prostate segmentation challenge. Since then, traditional ML models have been continually going down the leaderboard.

Ronneberger et al. designed the U-Net (figure 2) by dividing the network into two distinct parts: an encoding arm, progressively downsampling the input through convolutions and pooling similarly to the traditional CNN; and a decoding arm, completely symmetric to the encoding part, where at each upsampling step they bring the spatial coordinate information of the image from the opposite side of the network and use concatenation to place it together with the result of the up-convolution operation [11]. The model outputs a segmentation mask, one output for each pixel of the original image: the segmentation task was transformed into a pixel-wise classification task.

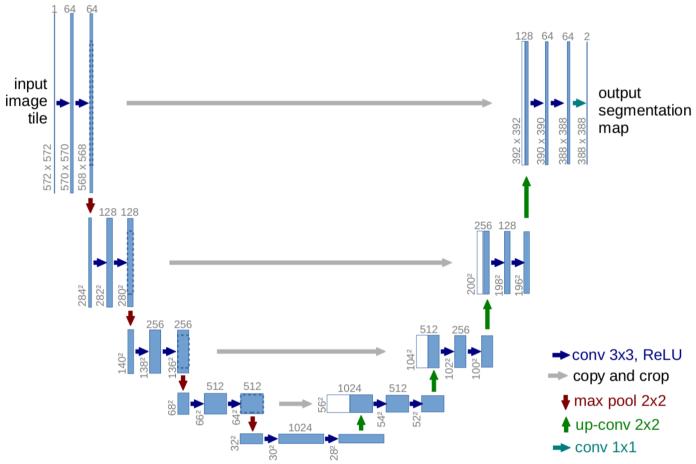

Fig. 2 Layout of the original U-net proposed in [11].

U-Net became widely used for medical imaging segmentation and several improvements were soon made. Cicek et al. created a version of U-net capable of using 3D inputs instead of 2D images [12]. Similarly, Milletari et al. proposed V-Net, a volumetric version of U-Net and incorporated the Dice coefficient into the loss function [13]. The advantage of having a 3D architecture is that instead of supplying a slice of a CT or MRI scan, one inputs the whole volume into the model, allowing for representation learning from all the data at once. Unlike a 2D model which loses information situated between slices and is incapable of inferring surface continuation, a 3D model can comprehend these details, which are especially useful at the top and bottom ends of each structure. This increases both accuracy and ease of use at the cost of computational capability.

3 Methods

A deep neural network was designed and trained on two datasets. The architecture of the network was based on the U-net and some additional features were added, described next.

The network has an encoder-decoder architecture, with a processing bottleneck - the deepest component of the network, where the image signal is the most encrypted - and high-resolution pathways, which convey the scale and general geometry information of the structures in the image to the decoder arm.

Figure 3 shows the general layout of the developed network, whose components are described in the following sections.

The model was implemented in Tensorflow using the Keras API, with the Python programming language.

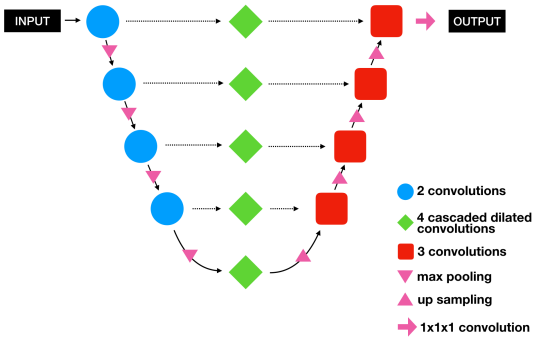

Fig. 3 Layout of the deep learning model as implemented.

3.1 Encoder arm

The first part of the network is composed of 4 levels, each with 2 convolutional layers and a max pooling layer.

Given that the input is three-dimensional, the convolutional operations are also performed in 3D such that unlike more typical 2D convolutions, the filters can be thought of as cubes instead of squares of trainable parameters. Padding with zeros was used where necessary to ensure that the output size was identical to the input volume at each convolutional layer. In this case, 3x3x3 filter sizes were used throughout, resulting in 27 weights per filter. The number of filters was doubled at each level such that the convolutional operations in the bottom level used 8 times as many filters as those in the first level. The number of filters was defined as a hyperparameter, to be tuned in the validation phase.

Additionally, there is one skip connection across both convolutional layers, mapping from the input of the first convolution to the output of the second, where they are joined by an element-wise addition operation; wherever the number of channels differs, an additional 1x1x1 convolution is applied in the skip connection which consolidates the number of channels in both tensors. This concept was taken from the ResNet architecture and is commonly called a residual connection [14]. This helps tackle the vanishing gradient problem typical of very deep neural networks, helps to better backpropagate the gradient to the first layers and offers some regularization thus reducing the tendency to overfit. Batch normalization was also implemented, as part of the full pre-activation residual connection block [15], as shown in Figure 4.

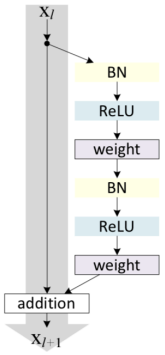

Fig. 4 A residual connection block as implemented. *BN* indicates batch normalization and *weight* indicates a convolutional layer (adapted from [15]).

The Max Pooling operation at the end of each level consists of a simple operation without trainable parameters, which uses a kernel size of 2x2x2 reducing therefore, the size of the input volume by 8 (decreasing to a half in each of the three dimensions). Therefore, given an input size of 256x256x128 (length x width x slices), the sizes of the tensors at the end of each of the encoding levels are 128x128x64, 64x64x32, 32x32x16 and 16x16x8. One can also see that the minimum size in any of the axes must be 16.

3.2 Decoder arm

Similarly to the encoding part of the network, the decoder arm is composed of 4 levels, and each has 3 convolutional layers. As shown in Figure 3, the first layer at each level of the decoder arm performs an Upsampling operation, in symmetry with the Max Pooling operation of the encoder arm. At this stage, the input tensor is increased by 2 at each dimension where each voxel becomes a 2x2x2 cube, increasing the total size 8-fold.

The resulting tensor is concatenated with the output of the skip pathway that brings in information from the encoder arm at the same level. This concatenation operation joins the highly encoded information of deeper layers with the spatial information of shallower layers, so that the network can more accurately make a decision on where to place the segmentation boundaries. This is similar to the original U-net, as proposed by Ronneberger et al. [11].

Afterwards, a set of three convolutional layers follows at each level, with one residual connection block as described above.

In the last decoder level at the top of the network as shown on Figure 3, there is a 1x1x1 convolutional layer, with a Softmax as activation function, which singles out the class with the highest prediction for each image voxel, outputting the final segmentation mask as predicted by the full network. In the training phase, this segmentation mask is compared with the ground-truth segmentation, i.e., the *labels*, and an error is calculated based on a loss function. Gradient descent is then used to backpropagate this error in order to update the network's trainable parameters and minimize the loss.

3.3 Dilated convolutions cascade

In order to enhance the segmentation ability and increase the receptive field of the network, some additional processing units were placed along the intermediate connections which bring information from the encoder to the decoder arms, as opposed to using bare skip connections.

In this case, dilated convolutional layers were employed, which process a larger part of the input thus gathering more information about the surrounding region around a given image point. The concept of dilation in convolutional layers corresponds to the number of pixels left between those whose information is processed; i.e., there are pixels whose values are not processed, making it so that although the size of the filters remain constant, by increasing the dilation rate we increase the receptive field (Figure 5).

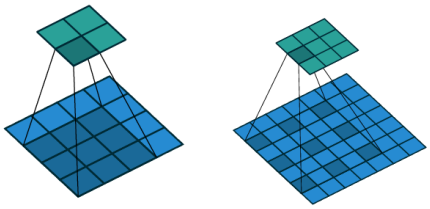

Fig. 5 Convolutions with dilation: (on the left) a standard convolution with a 3x3 filter size (dilation rate = 1) and (on the right) a convolution with dilation rate = 2 and the same filter size (from [16]).

The advantage of employing this technique, especially by doing it consecutively with different dilation rates, is to exploit information at different scales,

so as to better determine object boundaries [17]. Dilated convolutions can do this without increasing the number of parameters, hence without increasing complexity and training time.

In this work, a set of four dilated convolutions was employed in succession, i.e. a *cascade*, and their outputs joined by addition. At each layer the dilation rate is doubled, resulting in the following dilation rates for the four convolutional layers: 1, 2, 4, 8.

Across the whole cascade there is a residual connection that provides an alternative least-effort pathway where information can flow more easily, facilitating backpropagation, similarly to what is described in section 3.1. A dilated convolutions cascade is shown in Figure 6.

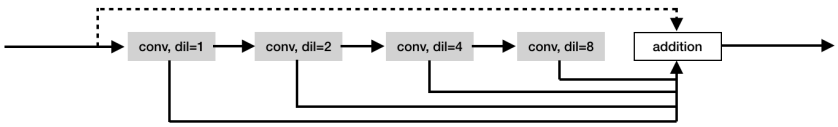

Fig. 6 A dilated convolutions cascade, as implemented in the proposed network: The outputs of four dilated convolutional layers are joined together and with a residual connection coming directly from the input.

Finally, the result is a powerful three-dimensional neural network, based on the U-net architecture with high-resolution skip pathways in the form of dilated convolutions cascades, that was trained end-to-end on an Nvidia DGX station with four Tesla V100 GPUs.

For the segmentation task, knowing that there is a large class imbalance - many more voxels belong to the background class than the target classes - instead of using Cross-entropy as the loss function during the training process, a loss function based on the Dice Coefficient was implemented, similar to the proposed in [13]. The optimization algorithm used was Adam [18], and the most appropriate learning rate was found by validation trials.

4 Results

The deep learning model described in the previous section was applied to two publicly-available datasets. For a complex task such as segmentation, it is hard to define a single metric that fully translates the capability of the model. Hence, a set of metrics were calculated and are shown. These include the Dice Similarity Coefficient as well as boundary-specific metrics such as the Average Boundary Distance (ABD) and the 95th percentile of the Hausdorff Distance (95% HD), both measured in mm, widely used in the literature for segmentation tasks [19].

The first dataset used is the Lung Computed Tomography Segmentation Challenge dataset (LCTSC) which contains thoracic studies of 36 patients, taken from different CT machines, with some variation in resolution and anatomical landmarks [20]. It was originally published as part of a challenge at AAPM 2017 [21]. The dataset is hosted by The Cancer Imaging Archive [22]. The ground-truth segmentations are divided into 6 classes: background, left and right lungs, heart, esophagus and spinal cord. In this work, 30 volumes were used as the training set and 6 for validation.

	DSC	Accuracy	Sensitivity	Specificity	ABD	95% HD
Spinal Cord	82.27	99.97	79.78	99.99	0.85	5.48
Left Lung	96.10	99.80	95.30	99.94	0.69	5.62
Right Lung	95.18	99.87	95.63	99.95	0.92	7.73
Heart	84.21	99.62	83.99	99.83	3.31	25.75
Esophagus	42.56	99.94	32.25	99.99	2.54	16.82

Table 1 Results for the Computed Tomography dataset.

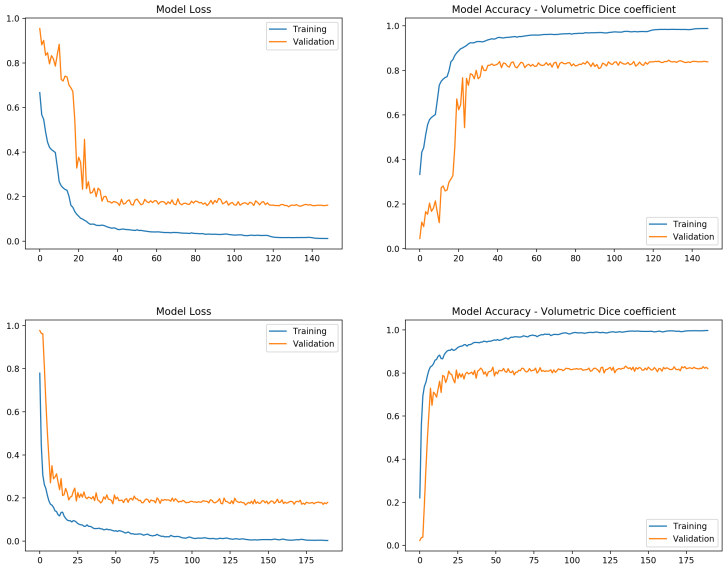

Fig. 7 The learning curves for the best model with the LCTSC (on the top) and PROMISE12 datasets (on the bottom): the loss of the model (on the left) and the Dice Coefficient (on the right) along the training epochs are despicted.

The volumes were composed of 512x512 images with a variable number of slices, from 115 to 279. In order to overcome the problem that all input samples must have the same input size, all images with fewer than 256 slices,

were padded with slices containing all zeros and the single study with more than 256 was cut at the bottom - those extra slices were not useful as they had no organ classes in the ground truth segmentations. The volumes were also downsized resulting in samples with size 256x256x128, which could be processed one at a time by the neural network as implemented.

Training was performed with some variations of the hyperparameters: learning rate and number of filters in the first level, with $3x10^{-4}$ and 26, respectively, producing the best results. Training took place over 150 epochs and the learning curves are shown on Figure 7. The Dice Coefficient achieved on the 6 samples of the validation data was 83.32, averaged over all patients and all organs. The best results were achieved in the lungs, with average DSC of 96.10 and the worst in the esophagus, with average DSC of 42.56, as shown in table 1.

The differences in the DSC for each organ type may be explained by the fact that larger organs, with well defined boundaries and large contrast differences such as the lungs are easier to contour than those with smaller dimensions and whose boundaries are so thin as to be nearly invisible in CT.

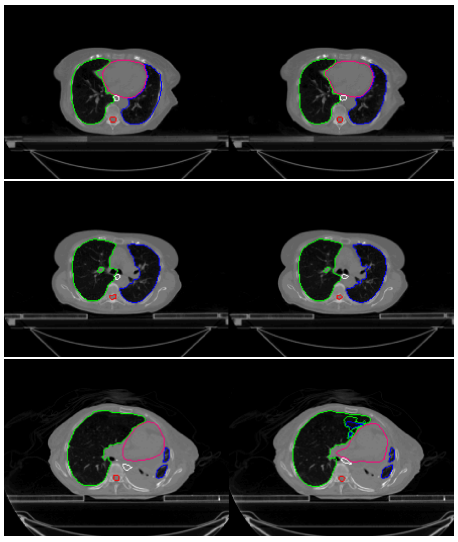

Fig. 8 Example comparisons of the ground-truth segmentations (on the left) with the model segmentations (on the right), for three different patients: On the top and middle rows, two accurate examples are shown; and on the bottom row, an example where the outcome was not so positive (in green: right lung, blue: left lung, red: spinal cord, white: esophagus, pink: heart).

In a visual comparison of the segmentations against the ground-truth, the model performs reasonably well, capturing the most significant landmarks with accuracy. The main limitations concern soft tissue structures, with low contrast, like heart and esophagus. However, as evidenced in the bottom row of Figure 8, even an easy organ can become an obstacle for computational mod-

	DSC	Accuracy	Sensitivity	Specificity	ABD	95% HD
Prostate	80.08	99.57	79.50	99.79	1.71	10.53

Table 2 Results for the Magnetic Resonance dataset.

els if it has considerable alterations; in this case, the patient has atelectasis, where the left lung is filled with liquid and solid tissue, making its radiographic signature very atypical and causing the other lung to expand to its side. Particularly, the model placed more emphasis on the side specificity than on the organ continuity.

The second dataset used was the Prostate Magnetic Resonance Image Segmentation (PROMISE12), which contains the most widely known dataset for prostate segmentation [2]. It was originally set up for the MICCAI conference in 2012 [23,24,25]. It is composed of a total of 80 studies, 50 publicly available for training and 30 kept private for benchmarking purposes. Therefore, in this work, only the 50 images in the training set were used, with 42 used for actual training and 8 for validation.

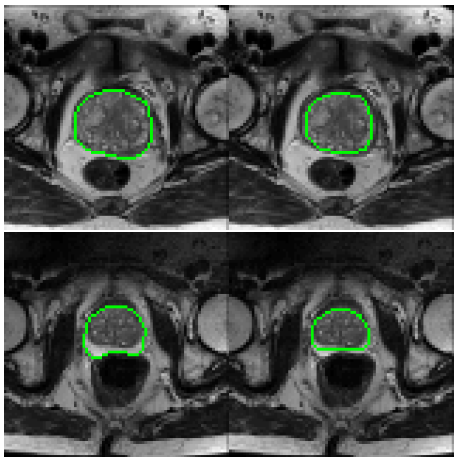

Fig. 9 Some example comparisons of the ground-truth segmentations (on the left) with the model segmentations (on the right), for two different patients.

In contrast with the CT dataset, only one organ was contoured, the prostate. The best validation DSC obtained was 80.08, after 180 epochs of training, shown in Figure 7. Additional evaluation metrics are shown in table 2. The best learning rate was found to be $3x10^{-3}$. A visual comparison with the ground-truth segmentations is shown in Figure 9. The model is able to identify the prostate gland, accurately separating it from the background, mostly avoiding pitfalls in the surrounding tissue with similar image features. On the bottom row, however, we can see that the model has difficulty in identifying the peripheral zone of the prostate (bottom of the segmentation), likely because it has similar intensity values to the surrounding tissue and a vanishing bound-

	DSC	Accuracy	Sensitivity	Specificity	ABD	95% HD
3D U-net	77.70	95.91	76.26	95.99	3.32	22.71
U^2-net	79.29	**99.86**	76.63	**99.97**	1.70	**11.04**
Our model	**80.06**	99.84	**77.39**	99.94	**1.66**	12.28

Table 3 Comparative results of different models on the Computed Tomography dataset.

	DSC	Accuracy	Sensitivity	Specificity	ABD	95% HD
3D U-net	79.43	99.21	**84.00**	99.50	2.49	11.38
U^2-net	**80.70**	99.23	76.12	**99.79**	2.26	12.13
Our model	80.08	**99.57**	79.50	**99.79**	**1.71**	**10.53**

Table 4 Comparative results of different models on the Magnetic Resonance dataset.

ary, while the boundary between the zones inside the prostate gland have more contrast.

4.1 Comparative experiments

In order to better assess the value of the proposed model, comparative experiments were performed with state of the art deep learning models which have the ability to perform segmentation on medical images: the 3D U-net proposed by Cicek et al. [12] and the U^2-net proposed by Qin et al. [26]. Both models were trained with scripts found in Github pages with the correct implementations provided by the authors [27, 28] and evaluated in the same manner as our model for fair comparisons.

For the CT dataset (LCTSC), the results are presented in Table 3, where the average across all organs (disregarding the background class) was taken for each metric, to allow for easier comparisons. Our model achieved the highest Dice score and also the lowest average boundary distance, although it was very close to U^2-net, which achieved a lower Hausdorff distance.

Regarding the PROMISE12 dataset, the results are very close among the three models, shown in Table 4. The highest volumetric Dice score was achieved by the U^2-net, whilst our model achieved the best result as to the boundary metrics.

Both the 3D U-net and the U^2-net use plain skip connections as the transfer of information from the encoding to the decoding arms of the network, whereas our model introduces the dilated convolutions cascade where additional processing of information is done, which can better capture fine edge details at a given level of abstraction and scale to be transferred to the decoding arm. This is likely the reason behind the better result on surface boundary metrics, as the cascade acts as a high-resolution pathway for edge information. On the other hand, the smaller U-net-like architecture at each level of the U^2-net can more accurately determine coarser organ shape features leading to a higher Dice score on the PROMISE12 validation set.

5 Conclusions

In summary, an image segmentation neural network was implemented, based on the original U-net and with some variations to improve its performance: fully three-dimensional architecture, residual connection blocks, skip connection high-resolution pathways through cascades of dilated convolutions, and a loss function based on the Dice Coefficient.

When trained on two very different datasets, LCTSC and PROMISE12, with CT and MRI images, respectively, the network achieved good results, as evidenced by the quantitative analysis and the visual comparisons against the ground-truth segmentations. This work verifies the versatility of deep learning models for segmentation, showing that the same architecture can be applied to medical imaging techniques with very different characteristics. No handcrafted feature extraction is required, facilitating the implementation. Furthermore, segmentation of multiple structures at once is viable and accurate with this method.

The additional network modules tailored to this model give it the ability to generate consistent segmentations, with smooth boundaries, reflecting the features of the underlying patient scan. It is able to use both the spatial information of the whole volume and the local features of the surrounding voxels to determine the presence of anatomic structures and organ boundaries. The model is particularly successful in large organs, and has some limitations in thin and long structures. Larger datasets could help to overcome these problems. However, more than just larger datasets, the requirement is for datasets which encompass a large variation of data, more representative of real world diversity.

As part of future work, we plan to apply this same network architecture to a curated dataset that we are collecting, with a large number of samples from a diverse population, in an attempt to better understand the limitations of this model and help improve its performance.

Compliance with Ethical Standards:

Funding: The authors would like to thank *Fundação para a Ciência e Tecnologia* (FCT) for the PhD grant (reference SFRH/BD/146887/2019) awarded to the first author, which this work is a part of.

Conflict of interest: The authors declare that they have no conflict of interest.

Ethical approval: This article does not contain any studies with human participants or animals performed by any of the authors.

References

1. Victor Spitzer, Michael J. Ackerman, Ann L. Scherzinger, and David Whitlock. The Visible Human Male: A Technical Report. *Journal of the American Medical Informatics Association*, 3(2):118–130, 03 1996.
2. Geert Litjens, Robert Toth, Wendy van de Ven, Caroline Hoeks, Sjoerd Kerkstra, Bram van Ginneken, Graham Vincent, Gwenael Guillard, Neil Birbeck, Jindang Zhang, Robin Strand, Filip Malmberg, Yangming Ou, Christos Davatzikos, Matthias Kirschner, Florian Jung, Jing Yuan, Wu Qiu, Qinquan Gao, Philip Eddie Edwards, Bianca Maan, Ferdinand van der Heijden, Soumya Ghose, Jhimli Mitra, Jason Dowling, Dean Barratt, Henkjan Huisman, and Anant Madabhushi. Evaluation of prostate segmentation algorithms for MRI: The PROMISE12 challenge. *Med Image Anal*, 18(2):359–373, 2014.
3. Mohammad Hesam Hesamian, Wenjing Jia, Xiangjian He, and Paul Kennedy. Deep Learning Techniques for Medical Image Segmentation: Achievements and Challenges. *J Digit Imaging*, 32(4):582–596, 2019.
4. H. P. Ng, S. H. Ong, K. W. C. Foong, P. S. Goh, and W. L. Nowinski. Medical image segmentation using k-means clustering and improved watershed algorithm. In *2006 IEEE Southwest Symposium on Image Analysis and Interpretation*, pages 61–65, 2006.
5. Eman Abdel-Maksoud, Mohammed Elmogy, and Rashid Al-Awadi. Brain tumor segmentation based on a hybrid clustering technique. *Egyptian Informatics Journal*, 16(1):71 – 81, 2015.
6. D. Jayadevappa, S. Srinivas Kumar, and D. S. Murty. Medical image segmentation algorithms using deformable models: A review. *IETE Technical Review*, 28(3):248–255, 2011.
7. Zhen Ma, Joao Manuel R.S. Tavares, Renato Natal Jorge, and T. Mascarenhas. A review of algorithms for medical image segmentation and their applications to the female pelvic cavity. *Computer Methods in Biomechanics and Biomedical Engineering*, 13(2):235–246, 2010. PMID: 19657801.
8. Chiranji Lal Chowdhary and D.P. Acharjya. Segmentation and feature extraction in medical imaging: A systematic review. *Procedia Computer Science*, 167:26 – 36, 2020. International Conference on Computational Intelligence and Data Science.
9. Yann LeCun, Yoshua Bengio, and Geoffrey Hinton. Deep learning. *Nature*, 521(7553):436–444, 2015.
10. Lequan Yu, Xin Yang, Hao Chen, Jing Qin, and Pheng-Ann Heng. Volumetric ConvNets with Mixed Residual Connections for Automated Prostate Segmentation from 3D MR Images. In *Thirty-First AAAI Conf Artif Intell*, pages 66–72, 2017.
11. Olaf Ronneberger, Philipp Fischer, and Thomas Brox. U-Net: Convolutional Networks for Biomedical Image Segmentation. In Nassir Navab, Joachim Hornegger, William M Wells, and Alejandro F Frangi, editors, *Med Image Comput Comput Interv – MICCAI 2015*, pages 234–241, Cham, 2015. Springer International Publishing.
12. Özgün Çiçek, Ahmed Abdulkadir, Soeren S. Lienkamp, Thomas Brox, and Olaf Ronneberger. 3D U-net: Learning dense volumetric segmentation from sparse annotation. *Lect Notes Comput Sci (including Subser Lect Notes Artif Intell Lect Notes Bioinformatics)*, 9901 LNCS:424–432, 2016.
13. Fausto Milletari, Nassir Navab, and Seyed Ahmad Ahmadi. V-Net: Fully convolutional neural networks for volumetric medical image segmentation. *Proc - 2016 4th Int Conf 3D Vision, 3DV 2016*, pages 565–571, 2016.
14. Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. *Proc IEEE Comput Soc Conf Comput Vis Pattern Recognit*, 2016-Decem:770–778, 2016.
15. Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Identity mappings in deep residual networks, 2016.
16. Vincent Dumoulin and Francesco Visin. A guide to convolution arithmetic for deep learning, 2016.
17. Sihang Zhou, Dong Nie, Ehsan Adeli, Jianping Yin, Jun Lian, and Dinggang Shen. High-Resolution Encoder-Decoder Networks for Low-Contrast Medical Image Segmentation. *IEEE Trans Image Process*, 29(X):461–475, 2019.
18. Diederik P. Kingma and Jimmy Ba. Adam: A method for stochastic optimization, 2014.

19. Varduhi Yeghiazaryan and Irina Voiculescu. Family of boundary overlap metrics for the evaluation of medical image segmentation. *J Med imaging (Bellingham, Wash)*, 5(1):15006, jan 2018.
20. Jinzhong Yang, Greg Sharp, Harini Veeraraghavan, Wouter van Elmpt, Andre Dekker, Tim Lustberg, and Mark Gooding. Data from lung ct segmentation challenge, 2017.
21. Jinzhong Yang, Harini Veeraraghavan, Samuel G. Armato III, Keyvan Farahani, Justin S. Kirby, Jayashree Kalpathy-Kramer, Wouter van Elmpt, Andre Dekker, Xiao Han, Xue Feng, Paul Aljabar, Bruno Oliveira, Brent van der Heyden, Leonid Zamdborg, Dao Lam, Mark Gooding, and Gregory C. Sharp. Autosegmentation for thoracic radiation treatment planning: A grand challenge at aapm 2017. *Medical Physics*, 45(10):4568–4581, 2018.
22. Kenneth Clark, Bruce Vendt, Kirk Smith, John Freymann, Justin Kirby, Paul Koppel, Stephen Moore, Stanley Phillips, David Maffitt, Michael Pringle, Lawrence Tarbox, and Fred Prior. The Cancer Imaging Archive (TCIA): Maintaining and Operating a Public Information Repository. *J Digit Imaging*, 26(6):1045–1057, 2013.
23. Qikui Zhu, Bo Du, and Pingkun Yan. Boundary-weighted Domain Adaptive Neural Network for Prostate MR Image Segmentation. *IEEE Trans Med Imaging*, pages 1–1, 2019.
24. Dong Nie, Li Wang, Yaozong Gao, Jun Lian, and Dinggang Shen. STRAINet: Spatially Varying sTochastic Residual AdversarIal Networks for MRI Pelvic Organ Segmentation. *IEEE Trans Neural Networks Learn Syst*, 30(5):1552–1564, 2019.
25. Haozhe Jia, Yong Xia, Yang Song, Donghao Zhang, Heng Huang, Yanning Zhang, and Weidong Cai. 3D APA-Net: 3D Adversarial Pyramid Anisotropic Convolutional Network for Prostate Segmentation in MR Images. *IEEE Trans Med Imaging*, PP(c):1–1, 2019.
26. Xuebin Qin, Zichen Zhang, Chenyang Huang, Masood Dehghan, Osmar R. Zaiane, and Martin Jagersand. U2-net: Going deeper with nested u-structure for salient object detection. *Pattern Recognition*, 106:107404, Oct 2020.
27. Yingkai (Kyle) Sha. Github repository for Keras Unet Collection, found at https://github.com/yingkaisha/keras-unet-collection.
28. Xuebin Qin and Zichen Zhang et al. Github repository for U2-Net, found at https://github.com/xuebinqin/U-2-Net, 2020.