Multi-organ Segmentation of Chest CT Images in Radiation Oncology: Comparison of Standard and Dilated UNet

Umair Javaid[1,2⋆], Damien Dasnoy[1], and John A. Lee[1,2]

1 - ICTEAM
Université Catholique de Louvain
Louvain-la-Neuve, Belgium
2 - IREC/MIRO
Université Catholique de Louvain
Brussels, Belgium
{umair.javaid,damien.dasnoy,john.lee}@uclouvain.be

Abstract. Automatic delineation of organs at risk (OAR) in Computed Tomography (CT) images is a crucial step for treatment planning in radiation oncology. However, manual delineation of organs is a challenging and time-consuming task subject to inter-observer variabilities. Automatic organ delineation has been relying on non-rigid registrations and atlases. However, lately deep learning appears as a strong competitor with specific architectures dedicated to image segmentation like UNet. In this paper, we first assess the standard UNet to delineate multiple organs in CT images. Second, we observe the effect of dilated convolutional layers in UNet to better capture the global context from the CT images and effectively learn the anatomy, which results in increased accuracy of organ delineation. We evaluate the performance of a standard UNet and a *dilated* UNet (with *dilated* convolutional layers) on four chest organs (esophagus, left lung, right lung, and spinal cord) from 29 lung image acquisitions and observe that *dilated* UNet delineates the soft tissues notably esophagus and spinal cord with higher accuracy than the standard UNet. We quantify the segmentation accuracy of both models by computing spatial overlap measures like Dice similarity coefficient, recall & precision, and Hausdorff distance. Compared to the standard UNet, *dilated* UNet yields the best Dice scores for soft organs whereas for lungs, both models have the same delineation accuracy: 0.84±0.07 vs 0.71±0.10 for esophagus, 0.99±0.01 vs 0.99±0.01 for left lung, 0.99±0.01 vs 0.99±0.01 for right lung and 0.91±0.05 vs 0.88±0.04 for spinal cord.

Keywords: multi-organ, segmentation, computed tomography.

1 Introduction

Radiation oncology treats cancer by irradiating solid tumors with photon or proton beams. Cancerous cells are slightly more sensitive to irradiation than healthy

⋆ Corresponding author.

ones and therefore daily targeted irradiation at low dose during several weeks maximizes the probability of local tumor control while minimizing undesired side effects. Medical images are used first as a map to localize the tumor (to be hit) and the organs at risk (to be avoided). Next, images also reveal physical properties of the tissues, which allow simulating deposition of energy by the treatment beams. Most steps of treatment planning are largely automated. However, a noticeable exception is delineation of the tumor and organs at risk (OAR) on the images. It can be seen as a ballistic problem with a trade-off between delivering maximum dosage to the target volume (TV) and minimum dosage to the OAR. Therefore, automatic and accurate delineation of OAR on Computed Tomography (CT) images is a crucial aspect of treatment planning in radiation oncology. Physicians and radiologists still perform this step manually where OAR and TVs are delineated in a slice-by-slice manner, this requires expertise and is a time-consuming task. Moreover, it is also prone to inter-observer variabilities. Therefore, the need to automate the organ delineation process that can result in efficient and better treatment planning.

Traditionally, automatic organ delineation has relied on non-rigid registrations and atlas-based methods [1,2,3]. Atlas-based segmentation uses only one or few examples, the atlas(es). Moreover, nonrigid registration considers simplistic deformation models through regularization of the deformation field, in some cases making it difficult to match the atlas with another patient anatomy. Several researchers analyzed hierarchical atlas registration and weighting scheme to generate target specific priors from an atlas database by combining aspects from multi-atlas registration and patch-based segmentation [4]. Others reported increased multi-organ segmentation accuracy by evaluating organ correlations and prior anatomical knowledge [5]. These approaches need prior information to capture anatomical context for the target organs and therefore, tend to establish anatomical similarities between images from different subjects. Such approaches are mostly handcrafted, therefore limited. They cannot be generalized sufficiently to capture a large number of variabilities of the deformable organs and hence, fail when the regions of interest (OAR or TV) undergo complex deformations.

Recently, deep learning methods have been shown to achieve state-of-the-art for a wider range of tasks in natural image processing such as classification, detection, and segmentation. Deep learning has become prevalent because of its remarkable capabilities to fully learn hierarchical organization of input data thanks to its multi-layered architectures. With convolutional neural networks (CNNs) achieving state-of-the-art performance in medical image segmentation [6], research has been ongoing to achieve better accuracy with higher spatial resolution. Fully convolutional network (FCN) is a variant of CNN that replaces fully connected layers by convolutions [7]. In a typical *supervised* learning setting for image segmentation, a FCN can learn the most suitable data associations between the input images and labels via layers of feature extractors using convolution operations, thus providing a segmentation map per available class in the labels.

In regards to biomedical image segmentation, UNet was introduced. It consists of a contracting path for capturing context and a symmetric expanding path that

enables precise localization of target regions in the image [8]. The architecture and the data augmentation of UNet allows efficient learning with superior generalization performance and is one of the most used network for medical image segmentation. Due to its unique architecture, many different approaches have been proposed using the same architectural choice. One study used patch-wise training strategy for bladder segmentation from CT urography in a slice-by-slice manner [9]. Other researchers evaluated several variations of deep ConvNets in the context of hierarchical, coarse-to-fine classification on image patches and regions i.e. superpixels for pancreas segmentation [10]. A third group proposed a hybrid densely connected UNet (H-DenseUNet), which consists of a 2D DenseUNet for efficiently extracting intra-slice features and a 3D counterpart for aggregating volumetric contexts for liver and tumor segmentation [11].

Delineating multiple organs in chest CT images can be formulated as a two-step, challenging task: 1) accurate localization of the organs and 2) delineation of the boundaries between organs and surrounding tissues. Deep learning networks achieve state-of-the-art in object detection and segmentation; their performance is, however, limited in segmenting soft tissues. Soft tissues show high variability in terms of shape, appearance, and relative spatial position with respect to different organs. Therefore they tend to have fuzzy boundaries and potentially same intensity values as the surrounding organs in the CT images that makes it difficult for the network to discriminate.

Even though the networks based upon the notion of UNet perform well for multi-organ segmentation, they are limited to a fixed kernel size i.e. the kernel size of the network does not increase nor decrease during operation. Researchers suggested dilated convolutions that allow multi-scale learning of the context. Dilated convolution preserves the spatial resolution of images i.e. the size of the feature map from one layer to another, which is achieved by increasing the field of view of the kernel [12]. Recently, one study used dilated convolutions to enlarge the network's receptive field size [13] whereas others analyzed *à trous*/dilated convolution for dense prediction tasks [14].

In this paper, we explore the discriminative capabilities of UNet to delineate multiple organs. We also evaluate the effect of dilated convolutions with UNet to capture the global context of the spatial relationship between different organs. We observe that compared to UNet, using dilated convolutions with UNet provides better delineation (*higher localization*) of the soft tissues such as esophagus and spinal cord whereas for lungs, both models performed considerably well and had comparable performance.

The remainder of the paper is organized as follows: Section 2 outlines the dataset used and the data pre-processing aspects. We present the different methods and models in Section 3. Experimental results are shown in Section 4 followed by discussion in Section 5. Conclusion and future perspectives are reported in Section 6.

2 Image data and Pre-processing

Dataset. Thirteen patients with histology-proven small cell lung cancer (SCLC, stage 2) were included in this study which is a retrospective use of their data [15]. Each patient was referred for curative intent chemoradiotherapy at UCLouvain University hospital Saint-Luc. Images were acquired at multiple time points during treatment. All scans were acquired on a Gemini TF PET/CT (Philips Medical Systems, Cleveland, OH). Patients were immobilized in treatment position with a thermoplastic thorax mask (Orfit Industries NV, Wijnegem, Belgium).

Each patient had multiple acquisitions where some acquisitions were excluded due to lack of manual annotations. A total of 29 image acquisitions with fully defined TVs and OARs were analyzed, where TVs and OARs were defined by the institutional guidelines and manually delineated by a single experienced radiation oncologist. The chest CT images and their corresponding reference OAR delineations, in particular esophagus, left lung, right lung, and spinal cord were used for this study.

Data pre-processing. To keep the data homogeneous across all patients, axial slices were selected from the 3D volume along the cranio-caudal direction around lungs followed by geometric re-sampling of the selected slices in terms of uniform pixel spacing (0.97 mm) and slice thickness (2 mm). Image intensity values of all scans were truncated to the range of $[-1000, 3000]$ HU (Hounsfield Unit) to remove possible artifacts in the high HU range. Finally, slices were downsampled to 256×256 and rescaled between 0 and 1 with 8 bits per pixel.

During training, input data was standardized by subtracting mean and dividing by the standard deviation where mean and standard deviation were calculated batch-wise per training iteration. A total of 900 images were considered as input. A test set of 50 slices was created by random selection of the input images, which was only used for evaluation purposes. A 20% validation split was used on the remaining input images, resulting in 680 training and 170 validation images respectively.

Data augmentation. Data augmentation techniques such as rotation, scaling, and translation were applied to the training images followed by random local elastic deformations. This resulted in local stretch and compression in images [16]. Such data augmentation was applied during training, which provided the network with new transformed examples, hence resulting in more effective learning per epoch. Rotation angle of 10° with a shift of 0.1 in height and width was used. Also, the images were zoomed with a factor of 0.01 where the magnitude of distortion for the elastic deformation was 500.

3 Methods

This study aims at exploiting the discriminative potential of deep learning networks for multiple organ delineation. We first briefly explain about UNet and

its use for organ delineation. Then we give an overview about dilated convolutions and their relevance for the delineation.

3.1 Network architecture and implementation

We used UNet, which is a multi-scale convolutional network. It consists of a contracting and an expanding path. The *contracting* step extracts features from data where each step of the contracting path consists of two convolutions with a 3×3 kernel, each followed by a ReLU activation along with 2×2 max-pooling with strides of two in each direction of the domain. On the other hand, the *expanding* path upsamples the feature maps and halves the number of feature channels. Each step in the expanding path uses unpooling, which is concatenated with the features in the contracting path by copy connections (shown in yellow in Fig. 1) during the same step.

The unpooling layers preserve the maximum activation locations which were extracted during the max-pooling step and re-use those locations to place the maximum activations back to their original spatial position. Fig. 1 shows the UNet architecture that we used. Images of the size 256×256 were used as input whereas the network produces a segmentation mask as output. The output mask is a pixel-by-pixel mask in which each pixel either belongs to one of the organs or the background. For this study, we had five classes: four organs (esophagus, left lung, right lung, and spinal cord) plus the background.

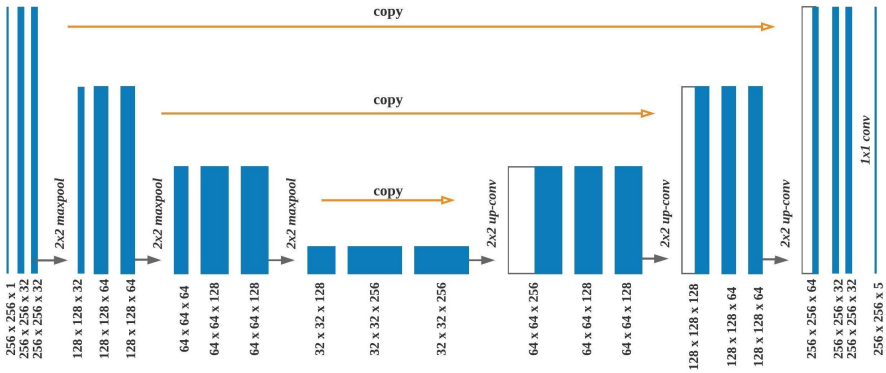


Fig. 1: All convolutons are 3×3 kernel size and ReLU as activation function except the last layer that is 1×1 and has *softmax* activation, where the output prediction involves five different classes.

3.2 Dilated convolutions

Dilated convolutions use sparse convolution kernel to enlarge the filters' field of view thus incorporating larger context for dense feature extraction. This is done

by deriving a larger filter from the original filter and dilating it by inserting zeros, thereby making it more efficient [17]. For a convolutional $k \times k$ size kernel, the size of the resulting dilated filter is $k_d \times k_d$ where $k_d = k + (k-1)(r-1)$, r is the dilation rate. Dilation rate is a parameter that introduces spacing between the kernel values. Compared to a normal convolution, a dilated convolution enables network operation on a coarser scale.

Dilated convolution when used in FCN potentially replaces the max-pooling operation while maintaining the field of view of the corresponding layer [18]. Spatial resolution of the feature map is preserved when the network traverses from one layer to another only by increasing the kernel's field of view. This process is applied iteratively to all layers having downsampling operations, thus the feature map in the output layer can maintain the same resolution as in the input layer. Dilated convolutions have previously been used in various contexts, e.g. signal processing [19], and image segmentation [14].

Precise organ delineation requires knowledge of the spatial relationship between different organs. In the innermost part of UNet, receptive field of the neurons is limited to a small subset of the image, thus limiting the information about relative organ positions. We therefore apply the dilated convolutions in the innermost bottleneck of UNet, where we replace the standard convolutions by dilated convolutions that we call *dilated* UNet. We used dilated convolution rates (2,4,8 and 16). Rest of the architecture remains the same for the two networks.

3.3 Quantitative analysis and evaluation

We use four metrics (three overlap measures and one contour-to-contour distance-based measure) to assess the similarity between predicted delineation per organ resulting from both networks and the ground-truth (manual organ delineation performed by the clinical expert). Dice similarity coefficient (DSC) measures the spatial overlap between ground-truth and predicted segmentations and therefore quantifies the match between two segmentation maps. For two binary masks A and B, DSC is the ratio of their intersection over their average cardinality

$$\mathrm{DSC(A,B)} = \frac{2|A \cap B|}{|A| + |B|}, \tag{1}$$

where $|A|$ counts the number of nonzero pixels in mask A. In the context of binary classification, recall/sensitivity is a measure of how good a classifier is at detecting the positives whereas precision is a measure of how many of the positively classified samples of data are relevant.

$$\text{recall/sensitivity} = \frac{\mathrm{TP}}{\mathrm{TP} + \mathrm{FN}}, \tag{2}$$

$$\text{precision} = \frac{\mathrm{TP}}{\mathrm{TP} + \mathrm{FP}}, \tag{3}$$

where TP, FP and FN stand for True Positive, False Positive, False Negative, and denote the number of pixels relevantly classified, the number of pixels irrelevantly classified, and the number of pixels that are ignored in the ground-truth

respectively. The Hausdorff distance (HD) is the maximal distance from a point in the first segmentation (predicted) to a nearest point in manual segmentation (ground-truth). Mathematically, it is defined as:

$$\mathrm{HD}(\mathrm{A},\mathrm{B}) = \max\{\max_{i\in A}\min_{j\in B} d(i,j), \max_{j\in B}\min_{i\in A} d(i,j)\}, \tag{4}$$

where $d(i,j)$ is the Euclidean distance between the centers of nonzero pixels i and j in A and B, respectively. HD is expressed in millimeters (mm). The above-mentioned metrics were computed on the predicted delineation for esophagus, left lung, right lung and spinal cord produced by both networks.

3.4 Network training

To train the networks, we used the standard pixel-wise unweighted cross-entropy as the loss function, which is given as

$$L(y,\hat{y}) = -\sum_{nc} y_{nc}\log\hat{y}_{nc}, \tag{5}$$

where n refers to pixels and c represents the classes, y_{nc} is the predicted delineation, and $\hat{y}_{nc}$ is the ground-truth delineation.

Networks were trained using the Adam optimizer with learning rate, LR = 0.001 and batch size = 32 for 3.96k iterations. Neither batch normalization nor dropout was used for the simulations. Training each network took almost 3 hours where training was done using two Titan X (Pascal) GPUs. Keras [20] was used to implement both networks.

4 Results

We report hereby different results of experiments that were conducted to quantify the performance of both networks in terms of multiple organ delineation. We used spatial overlap measures such as Dice score, precision and recall to evaluate the predicted delineation masks as given in Table 1 and 2 respectively. It can be seen from the tables that *dilated* UNet yields higher Dice scores for both esophagus and spinal cord, whereas both networks perform almost equally good for delineation of lungs. Some entries in the tables appear bold for a quick comparison between the two tables and indicate the performance gain between the two networks.

Low recall signifies that the network is not able to identify the whole organ, thus missing some spatial details and resulting in coarse segmentation (*spinal cord with UNet*, Table 1). Low precision means that the network encompasses wrong details as part of the target organ, resulting in fragmented segmentation (*esophagus with UNet*, Table 1). We observe that *dilated* UNet results in higher recall and precision compared to UNet, highlighting more effective learning of the spatial context between organs.

Figure. 2 shows the box plots of Dice score per organ evaluated for the predictions from both networks. It can be observed that *dilated* UNet offers higher Dice score than UNet especially for the esophagus and spinal cord.

Table 1: Per organ quantitative measures computed on UNet predictions.

Organs	Dice	TPR	FPR	Precision	Recall
Esophagus	**0.71±0.10**	0.78±0.13	0.32±0.15	**0.68 ±0.15**	**0.78 ±0.13**
Left lung	0.99±0.01	0.99±0.01	0.02±0.01	0.98±0.01	0.99±0.01
Right lung	0.99±0.01	0.99±0.01	0.02±0.01	0.98±0.01	0.99±0.01
Spinal cord	**0.88±0.04**	0.86±0.09	0.09±0.07	**0.91±0.07**	**0.86±0.09**

Table 2: Per organ quantitative measures computed on *dilated* UNet predictions.

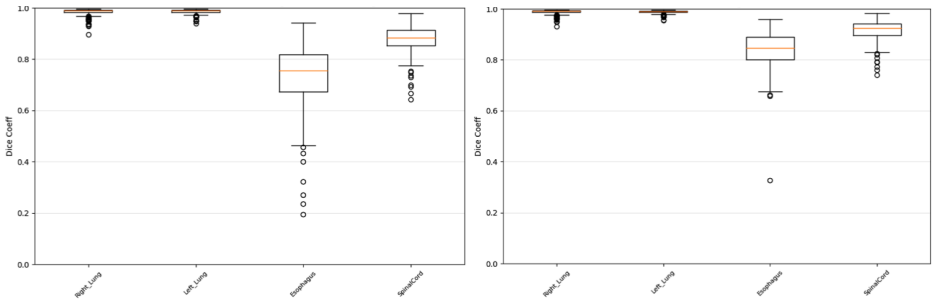
Organs	Dice	TPR	FPR	Precision	Recall
Esophagus	**0.84±0.07**	0.82±0.09	0.15 ±0.09	**0.85 ±0.09**	**0.82 ±0.09**
Left lung	0.99±0.01	0.99±0.00	0.02±0.01	0.98±0.01	0.99±0.00
Right lung	0.99±0.01	0.99±0.00	0.02±0.01	0.98±0.01	0.99±0.00
Spinal cord	**0.91±0.05**	0.92±0.06	0.10 ±0.09	**0.90±0.09**	**0.92±0.06**

We checked the Hausdorff distance; (HD) to compute the distance between the predicted delineation and ground-truth. We report the results in Fig. 3. It can be seen that *dilated* UNet offers a significantly better contour consistency compared to standard UNet. As per *dilated* UNet delineation, both lungs lie well below 5 mm whereas esophagus and spinal cord lie below 4 and 3 mm respectively. HD can be considered as finding the worst-case scenario in the delineation. As organs undergo anatomical variations, the mentioned numbers (HD distances per organ) are practical in clinical applications.

Finally, we report the predicted delineations from both networks in Fig. 4. Each image is plotted against the ground-truth delineations where light colors correspond to ground-truth and dark colors correspond to the predicted delineations. It can be observed that both networks capture the context and delineate all the target organs. However, UNet cannot precisely localize the organ boundaries especially for esophagus and spinal cord (low recall and precision scenario). On the other hand, *dilated* UNet incorporates better spatial context and is therefore able to localize the target organs with higher accuracy than UNet.

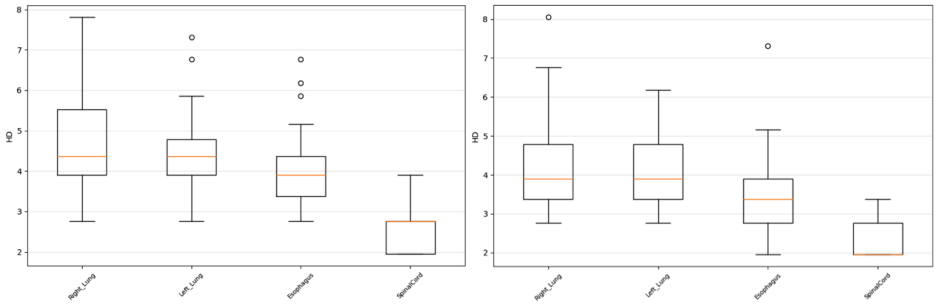
5 Discussion

Automatic delineation of multiple organs on CT images plays an important role in treatment planning in radiation oncology. As cancerous cells (target volumes - TVs) are treated by irradiation, the purpose of radiotherapy is to provide maximum dosage to the TV and minimum dosage to the OAR. Delineation helps medical staff in the diagnostic process by providing the contours for OAR and TV. In this paper, we try to automate the multi-organ delineation step by exploring the discriminative learning capabilities of deep learning networks. In particular, first we use one of the most commonly used network i.e. UNet for delineation and later, we modify the same network by replacing the standard



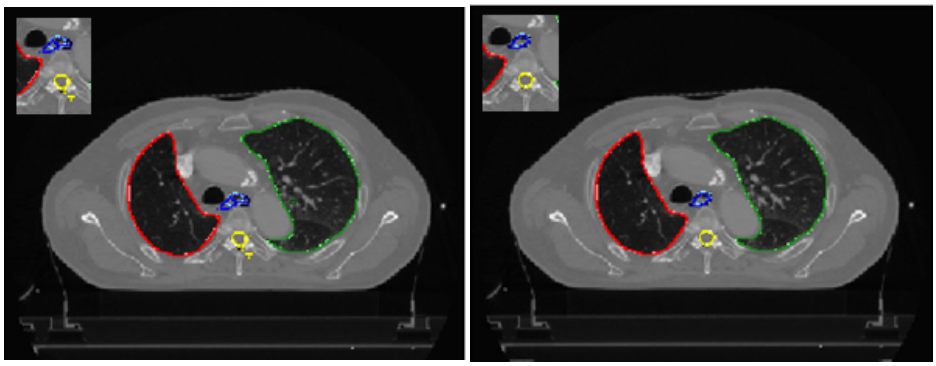
(a) UNet (b) Dilated UNet

Fig. 2: Box plots showing dice score per organ.



(a) UNet (b) Dilated UNet

Fig. 3: Box plots showing Hausdorff distance per organ in mm.



(a) UNet (b) Dilated UNet

Fig. 4: Organ delineation (Ground-truth, GT vs Predicted Mask, PM), light color is for GT whereas dark color is for PM (best viewed in color).

convolutions with dilated convolutions and observe the increased performance in terms of delineation by the modified network that we call *dilated* UNet.

Information about the spatial relationship between organs is needed for accurate delineation of organs. Downsampling operation is a common practice to reduce the spatial dimensions of image to provide a wider receptive field to the neurons. However, downsampling operation is costly as it results in additional network parameters. In this study, instead of using additional downsampling operations, we used dilated convolutions in the innermost bottleneck of UNet to increase the receptive fields of the network. Dilated convolutions allow an exponential growth of receptive field with a linear increase in the number of parameters. We used dilation rates of 2, 4, 8 and 16 to incorporate more spatial context due to wider kernel sizes of the convolutions. We observed accurate localization of the organs by *dilated* UNet, thus highlighting the property of dilated convolutions to capture the spatial relationship between organs.

Medical practitioners usually perform manual delineation that is used as a ground-truth for data analysis. There is a risk that the manual annotation might be done in a hurry or lack accuracy. For example, a physician might find some organs less relevant for the radiation treatment planning, therefore such organs might be roughly or not delineated at all. Interpolation of contours across slices also occurs frequently in manual delineation. Therefore, in order to benchmark the deep learning performance versus human-level accuracy, we need an accurately annotated ground-truth that offers effective learning of the anatomical changes and hence, help the doctors to devise better and more precise treatment planning. Relying on retrospective images taken from clinical routine might be suboptimal, compared to images that could be prospectively collected and delineated specifically. While this has been feasible for atlases with only one or few images, the cost in time and money for deep learning, where many examples are necessary, might turn out much too high.

Having a valid ground-truth, we also need to have a standardized data augmentation. As annotated medical data is limited and deep learning networks are data-demanding, data augmentation techniques are a common practice to increase the training dataset. We need to introduce a controlled amount of variability that is consistent with actual variations. For example, flipping a subject does not make sense as the anatomy is not symmetrical and it can degrade the network learning.

There should also be a way to include inter-user variability into the deep learning networks. It is a usual practice to augment the images but we should also try to augment the contours. By doing so, we will potentially end up with multiple contours per organ and hence, multiple ground-truths, which could be a surrogate for the actual inter-observer variabilities.

Finally, uniformity of the medical data acquired from different hospitals should be considered. Having a standardized database can result in faster and efficient data analyses, allowing better insights into the anatomical changes.

6 Conclusion and Future Perspectives

In this paper, we assessed UNet for multiple organ delineation in chest CT images. We also explored the effect of dilated convolutional layers in UNet (*dilated* UNet) instead of standard convolutions and observed an increased performance in soft tissue delineation, notably the esophagus and spinal cord. We evaluated an alternative to deep networks as the common notion indicates that for better learning, the network should be deep enough. Here, we benchmark the performance of UNet without increasing its depth. We observed that using dilated convolutions with UNet resulted in improved localization for multiple organs delineation, especially for esophagus and spinal cord.

Having observed the effectiveness of *dilated* UNet to incorporate better global context in the images, we aim to extend our study to 3D and to higher number of organs. We also intend to try our *dilated* UNet on other datasets such as pelvic region and/or head and neck delineation.

Acknowledgments

Umair Javaid is funded by Fonds de la Recherche Scientifique - FNRS, Télévie grant no. 7.4625.16. Damien Dasnoy is a Research Fellow of FNRS. John A. Lee is a Senior Research Associate with the Belgian FNRS. We thank UCLouvain University hospital Saint-Luc for providing the data. We also thank NVIDIA Corporation for providing Titan X (Pascal) GPUs.

References

1. Tiezhi Zhang, Yuwei Chi, Elisa Meldolesi, and Di Yan. Automatic delineation of on-line head-and-neck computed tomography images: toward on-line adaptive radiotherapy. *International Journal of Radiation Oncology• Biology• Physics*, 68(2):522–530, 2007.
2. Subrahmanyam Gorthi, Valérie Duay, Nawal Houhou, Meritxell Bach Cuadra, Ulrike Schick, Minerva Becker, Abdelkarim S Allal, and Jean-Philippe Thiran. Segmentation of head and neck lymph node regions for radiotherapy planning using active contour-based atlas registration. *IEEE Journal of selected topics in signal processing*, 3(1):135–147, 2009.
3. Jose Dolz, HA Kirişli, T Fechter, S Karnitzki, O Oehlke, U Nestle, Maximilien Vermandel, and L Massoptier. Interactive contour delineation of organs at risk in radiotherapy: clinical evaluation on NSCLC patients. *Medical physics*, 43(5):2569–2580, 2016.
4. Robin Wolz, Chengwen Chu, Kazunari Misawa, Michitaka Fujiwara, Kensaku Mori, and Daniel Rueckert. Automated abdominal multi-organ segmentation with subject-specific atlas generation. *IEEE transactions on medical imaging*, 32(9):1723–1730, 2013.
5. Toshiyuki Okada, Marius George Linguraru, Masatoshi Hori, Ronald M Summers, Noriyuki Tomiyama, and Yoshinobu Sato. Abdominal multi-organ segmentation from CT images using conditional shape–location and unsupervised intensity priors. *Medical image analysis*, 26(1):1–18, 2015.

6. Geert Litjens, Thijs Kooi, Babak Ehteshami Bejnordi, Arnaud Arindra Adiyoso Setio, Francesco Ciompi, Mohsen Ghafoorian, Jeroen AWM van der Laak, Bram van Ginneken, and Clara I Sánchez. A survey on deep learning in medical image analysis. *Medical image analysis*, 42:60–88, 2017.
7. Jonathan Long, Evan Shelhamer, and Trevor Darrell. Fully convolutional networks for semantic segmentation. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 3431–3440, 2015.
8. Olaf Ronneberger, Philipp Fischer, and Thomas Brox. U-net: Convolutional networks for biomedical image segmentation. In *International Conference on Medical image computing and computer-assisted intervention*, pages 234–241. Springer, 2015.
9. Kenny H Cha, Lubomir Hadjiiski, Ravi K Samala, Heang-Ping Chan, Elaine M Caoili, and Richard H Cohan. Urinary bladder segmentation in CT urography using deep-learning convolutional neural network and level sets. *Medical physics*, 43(4):1882–1896, 2016.
10. Holger R Roth, Le Lu, Amal Farag, Hoo-Chang Shin, Jiamin Liu, Evrim B Turkbey, and Ronald M Summers. Deeporgan: Multi-level deep convolutional networks for automated pancreas segmentation. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 556–564. Springer, 2015.
11. Xiaomeng Li, Hao Chen, Xiaojuan Qi, Qi Dou, Chi-Wing Fu, and Pheng Ann Heng. H-DenseUNet: Hybrid densely connected UNet for liver and liver tumor segmentation from ct volumes. *arXiv preprint arXiv:1709.07330*, 2017.
12. Fisher Yu and Vladlen Koltun. Multi-scale context aggregation by dilated convolutions. *arXiv preprint arXiv:1511.07122*, 2015.
13. Eli Gibson, Francesco Giganti, Yipeng Hu, Ester Bonmati, Steve Bandula, Kurinchi Gurusamy, Brian R Davidson, Stephen P Pereira, Matthew J Clarkson, and Dean C Barratt. Towards image-guided pancreas and biliary endoscopy: Automatic multi-organ segmentation on abdominal CT with dense dilated networks. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 728–736. Springer, 2017.
14. Liang-Chieh Chen, George Papandreou, Iasonas Kokkinos, Kevin Murphy, and Alan L Yuille. Deeplab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected CRFs. *IEEE transactions on pattern analysis and machine intelligence*, 40(4):834–848, 2018.
15. Dario Di Perri, John A Lee, Anne Bol, Francois-Xavier Hanin, Guillaume Janssens, Daniel Labar, Annie Robert, Edmond Sterpin, and Xavier Geets. Evolution of [18f] fluorodeoxyglucose and [18f] fluoroazomycin arabinoside PET uptake distributions in lung tumours during radiation therapy. *Acta Oncologica*, 56(4):516–524, 2017.
16. Patrice Y Simard, David Steinkraus, John C Platt, et al. Best practices for convolutional neural networks applied to visual document analysis. In *ICDAR*, volume 3, pages 958–962, 2003.
17. Pierre Dutilleux. An implementation of the "algorithme à trous" to compute the wavelet transform. In *Wavelets*, pages 298–304. Springer, 1990.
18. Panqu Wang, Pengfei Chen, Ye Yuan, Ding Liu, Zehua Huang, Xiaodi Hou, and Garrison Cottrell. Understanding convolution for semantic segmentation. *arXiv preprint arXiv:1702.08502*, 2017.
19. Aaron Van Den Oord, Sander Dieleman, Heiga Zen, Karen Simonyan, Oriol Vinyals, Alex Graves, Nal Kalchbrenner, Andrew Senior, and Koray Kavukcuoglu. Wavenet: A generative model for raw audio. *arXiv preprint arXiv:1609.03499*, 2016.
20. François Chollet et al. Keras, 2015.