Segmentation of Head and Neck Tumours Using Modified U-net

Baixiang Zhao, John Soraghan, and
Gaetano Di Caterina,
Centre for Signal and Image processing, Department of electronics and electrical enginering
University of Strathclyde,
Glasgow,United Kingdom
baixiang.zhao@strath.ac.uk,
j.soraghan@strath.ac.uk,

Derek Grose
Beatson West of Scotland Cancer Centre

Glasgow,United Kingdom
Derek.Grose@ggc.scot.nhs.uk

***Abstract*— A new neural network for automatic head and neck cancer (HNC) segmentation from magnetic resonance imaging (MRI) is presented. The proposed neural network is based on U-net, which combines features from different resolutions to achieve end-to-end locating and segmentation of medical images. In this work, the dilated convolution is introduced into U-net, to obtain larger receptive field so that extract multi-scale features. Also, this network uses Dice loss to reduce the imbalance between classes. The proposed algorithm is trained and tested on real MRI data. The cross-validation results show that the new network outperformed the original U-net by 5% (Dice score) on head and neck tumour segmentation.**

Keywords— MRI data, Head and neck cancer, U-net, dilated convolution, semantic segmentation

I. INTRODUCTION

According to the World Health Organisation [1], approximately 8.8 million people worldwide died from cancer in 2015. Radiotherapy, along with surgery, provides the main option for curative treatment. Radiotherapy planning is a complicated and lengthy process requiring detailed defining of complex cancer regions. This area is referred to as the Gross Tumour Volume (GTV). Definition of this region is fundamental to accurate and effective radiation treatment planning. Development of automated delineation methods can reduce inter and intra variabilities of manual tumour delineation, and provide objective and reliable assistance to clinical oncologists to reduce work load and improve radiation treatment [2].

Fig.1 (a) shows a T1 weighted gadolinium-enhanced head and neck MR image with tongue base tumours. It is known that the tumour region has fuzzy boundaries and it is not significantly distinct from neighbour tissues. Also, in head and neck region, there are regions and tissues has similar features (intensities, locations) with tumours, such as lymph nodes; these may produce false positives. Furthermore as seen in Fig.1 (b) artefacts of MRI data, such as uneven illumination, are obvious. All these make automatic tumour segmentation a very challenging task.

A variety of algorithms have been proposed for head and neck cancer or tissue segmentation, such as atlas-based techniques [3], training-based approaches [4, 5], and Deformable model [6-8]. However, these methods cannot efficiently solve the automatic segmentation challenge. The work using deformable models relies on quality of initialisation; while in [3], the atlas-based and [4] the training-based approaches relies on an atlas or large amount of labelled data. Also, from the review of deep learning research on medical image [9], currently no efficient deep learning approach is applied on head and neck cancer segmentation.

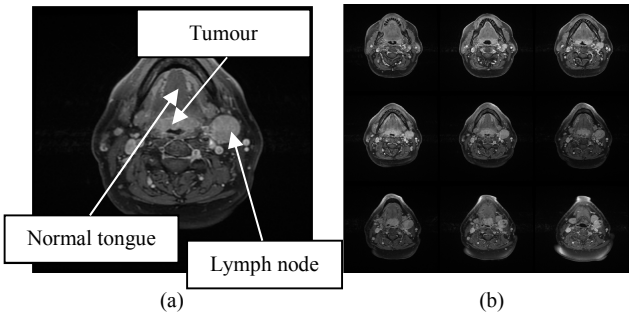


Fig. 1. (a) A T1-weighted Gadolinium-enhanced head and neck MR image example with cancerous lymph nodes; (b) A image serires from one MRI dataset.

This paper presents an end-to-end algorithm to segment head and neck tumours from MRI data. The challenges of this work include segmenting tumour regions with fuzzy boundaries, non-uniform intensities, and avoiding adjacent anatomical structures; also the discrimination of true targets from similar tissues and regions; and finally, train a deep neural network with limited numbers of data. It is essential to locate the position of cancer region and extract the exact cancer area. This algorithm is validated on real MRI data from the Beatson West of Scotland Cancer Centre, in Glasgow.

The remainder of the paper is organised as follows. Section 2 describes the new deep neural network for HNC segmentation. Section 3 demonstrates the experimental results on real MRI datasets. The last section summaries the paper.

II. U-NET WITH BROADER VIEW

The proposed head and neck tumour segmentation network is shown in Fig. 2. This section will introduce the proposed network from three aspects: a) U-net architecture, b) Dilated convolution, c) Dice loss.

A. U-net: biomedical image segmentation network

Convolutional neural networks (CNN) are powerful on feature extraction. By combining convolutional layers and fully connected layers, CNN shows its abilities on image recognition and classification [10]. Recently, many CNN

*Research supported by Beatson Cancer Charity.

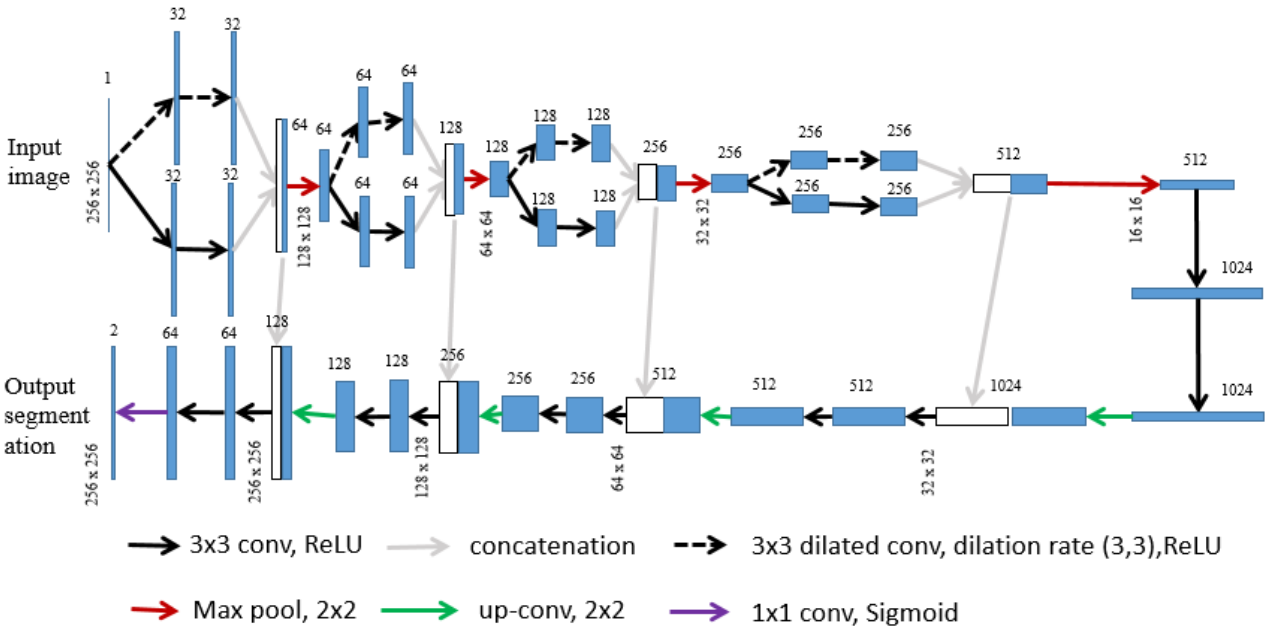


Fig. 2. The architecture of proposed neural network. Each box represents a multi-channel feature map, and the white and blue box are used together to show the merge (concatenation) of feature maps. The number of channels is denoted on top of the box. The vertical text 'mxn' means the xy size of feature map. The arrows represent the different operations.

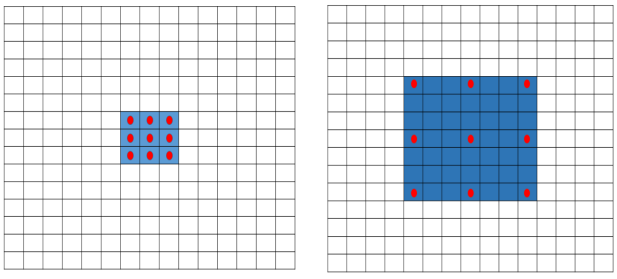
based networks gave the state-of-the-art performances on image classifications [11, 12]. Different from image classifications, the semantic segmentation tasks require detection of target locations and classification of every single pixel in images. In [13], a 'fully convolutional' networks (FCN) was built for end-to-end, pixels-to-pixels semantic segmentation. The FCN used skip connections between contracting (down-sampling) and expansive (up-sampling) paths, so that combined semantic information from deep, coarse layers with appearance information from shallow, fine layers to improve the segmentation performance [13]. The U-net [14] utilised structure of FCN, while the deep and shallow features are combined using concatenation instead of addition. U-net has been widely used in medical image segmentation, such as [15,16].

This work proposes a neural work architecture similar with U-net. As shown in Fig. 2, there are two major paths in the proposed architecture.: one is the contracting path at the top row from left to right, another one is the expansive path at the bottom row from right to left. In the contracting path, the xy sizes of feature maps are decreasing and the channels (z size) of feature maps are increasing, here features from different scales are extracted. In the expansive path, feature maps are up-sampled to the original resolution of input image to finish pixels-to-pixels segmentation. There are concatenations in contracting path, which merges the features extracted from normal convolution and dilated convolution. Also, concatenations are used between contracting and expansive paths to feed high resolution features into up-sampled output so that improve the precision.

The proposed network has 31 convolutional layers in total, this model can be further compressed in future work. This work aims to introduce multi-scale feature extraction into U-net by using dilated convolution, so that improve the performance of U-net on head and neck tumour segmentation. The dilated convolution will be introduced in following section.

B. *Dilated convolution for larger context view*

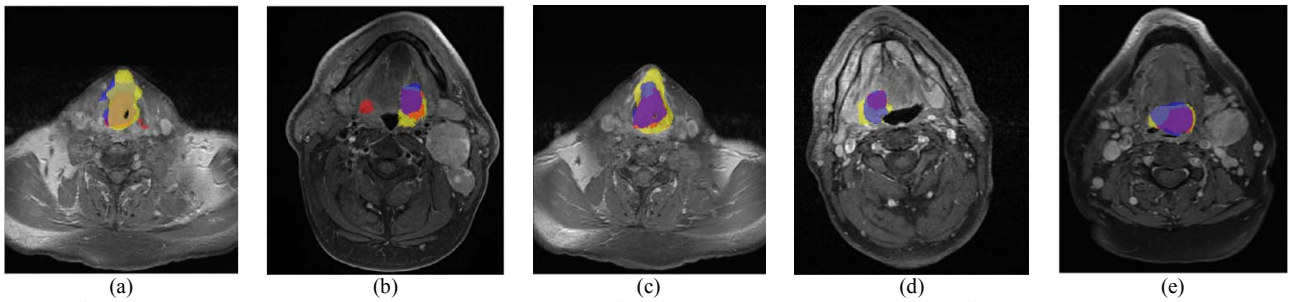
In deep neural network, one significant advantage of convolution and pooling is encoding regional and multi-scale information. While there are drawbacks, to access a larger receptive field to obtain non-local features, it will take many layers by using small size convolutions, or take lots of parameters by using large size convolutions. Also, even pooling layer could help extract the multi-scale information extraction, it will reduce the resolution of features. Thus, in [17] the dilated convolution was introduced to aggregate multi-scale contextual information without losing resolution or coverage.



(a) (b)

Fig. 3. Comparation between a normal 3x3 convolution kernel and a 3x3 dilated convolution kernel with dilation rate of (3,3). The weights of kernel are on red dots, and the blue regions represent the receptive field. In (a), the 3x3 kernel has a receptive field of 3x3. In (b), the dilated kernel can have receptive field of 7x7.

In Fig.3 the difference between normal convolution and dilated convolution is shown. It can be seen that using same numbers of parameters (e.g. 3x3), the dilated convolution can have broader view of context information in images. As demonstrated in Fig.2, in the contracting path of proposed network, dilated convolutions and normal convolutions are used simultaneously to get multi-scale features. And the extracted multi-scale features are combined using concatenations to improve the performance of network.



(a) (b) (c) (d) (e)

Fig.4. Visualization of segmenation results. (a)-(e) are 2D MRI slices from different patients. In each image (slice) yellow area represents true label (consensus manual delineation), red area represents segmentation from U-net, and blue area represents results from proposed work.

C. *Dice loss for classes balance*

There are several kinds of loss functions used in deep neural network training. A loss function takes all classes with equal weights (such as cross-entropy) is not suitable for head and neck tumour segmentation task. This is because that in head and neck image, the tumour region (foreground) is small compared to entire image but it has most importance; while background takes majority of areas in a head and neck image. Thus a Dice loss is introduced in [18] to solve the imbalance between foreground and background classes. The Dice loss takes the idea of Dice coefficient score (DSC) [19], DSC measures the similarity between two samples. The DSC between prediction(segmentation) and true label can be,

$$D = \frac{2 \times P \cap T}{P + T} \tag{1}$$

where P denotes predicted label, and T denotes true label. And in this way, the Dice loss will be,

$$L_D = 1 - D \tag{2}$$

The DSC attaches more importance to true positive (foreground) while give less to true negative (background), thus the imbalance between tumour and non-tumour classes is minimized.

III. Experimental results

The new algorithms were implemented in Matlab, running on PC with 16G RAM, 3.2GHz Intel(R) Core(TM) i7-8700 CPU, and a NVIDIA GTX 1070 GPU . Experiments were on real MRI datasets from Beatson west Scotland cancer centre. Totally 163 images (2D slices with about 3mm distance between slices) from 17 patients were used in this work. The true labels are consensus manual delineation provided by clinicians from Beatson.

This work is implemented with Keras using Adam [20] for optimization. Data augmentation including rotation (0.2 range), shift on horizontal (0.05) and vertical (0.05) direction, shear (0.05), zoom (0.05), and flip (horizontal) are used to improve the efficiency of usage of annotated data. Also, batch normalizations [21] are used to improve the learning process. The 163 images are split into three subsets with 0.3, 0.3, and 0.4 proportion, noted as $\{S_1, S_2, S_3\}$. Then the proposed network and U-net are measured with 3-fold cross-validation, in validation 1 S_1 and S_2 are used for training and S_3 is for testing, in validation 2 S_2 and S_3 are for training and S_1 is for testing, and finally in validation 3 S_1 S_3 are for training and S_2 is for testing. The average DSC of the cross-validation is shown in following table,

TABLE 1 HNC SEGMENTATION PERFORMANCE COMPARISON

Methods	**3-fold cross-validation**		
	Validation 1	***Validation 2***	***Validation 3***
U-net	0.6230	0.5576	0.5954
Proposed	0.6735	0.6076	0.6510

The results show that on HNC segmentation the proposed approach achieves about 0.644 dice score, this is about 0.05 higher than original U-net. And some segmentation comparisons are displayed in Fig.4. As here U-net is also trained with Dice loss to have better performance on these imbalance data, and U-net and proposed method use same augmentation setting, thus the improvement of DSC should be majorly from the dilated convolution layers added in contracting path. The bad results may come from limited numbers of data and serious artefacts (uneven illuminance, etc.) in parts of data. In this work, the data and labels are separate 2D slices with about 3 mm gaps between slices, thus by far only 2D work is conducted. The following work for 3D volume extraction can be achieved in two approaches: interpolate the 2D results (contours) to generate 3D meshes; alternatively, interpolate data and labels first, and design 3D convolutional networks. These will be explored in the future work and compared with 2D work.

IV. Conclusion

This paper presented a new deep neural network for semantic segmentation of head and neck tumour from T1-weighted Gadolinium-enhanced head and neck MR image. The proposed method was shown to work well on real MRI datasets. The results on real data show that this network can segment most tumour regions and outperformed U-net on HNC segmentation.

In the future, this method will be tested on more MRI datasets. The proposed network architecture will be further modified to increase the DSC. Also, more data augmentation methods will be explored to improve the network's performance on small numbers of data. And, the pre-processing of medical image artefacts can be applied to improve performance. Finally, the work could be extended to 3D in the future.

Acknowledgment

The authors would like to acknowledge grant from Beatson Cancer Charity for their financial support with this study.

References

[1] (March 03). *Head and Neck Cancers*. Available: https://www.cancer.gov/types/head-and-neck/head-neck-fact-sheet#r28

[2] T. Doshi, C. Wilson, C. Paterson, C. Lamb, A. James, K. MacKenzie*, et al.*, "Validation of a Magnetic Resonance Imaging-based Auto-contouring Software Tool for Gross Tumour Delineation in Head and Neck Cancer Radiotherapy Planning," *Clinical Oncology,* vol. 29, pp. 60-67, 2017/01/01/ 2017.

[3] X. Han, M. S. Hoogeman, P. C. Levendag, L. S. Hibbard, D. N. Teguh, P. Voet*, et al.*, "Atlas-based auto-segmentation of head and neck CT images," in *International Conference on Medical Image Computing and Computer-assisted Intervention*, 2008, pp. 434-441.

[4] Y. Zhang, M. T. Ying, L. Yang, A. T. Ahuja, and D. Z. Chen, "Coarse-to-fine stacked fully convolutional nets for lymph node segmentation in ultrasound images," in *2016 IEEE International Conference on Bioinformatics and Biomedicine (BIBM)*, 2016, pp. 443-448.

[5] K. Fritscher, P. Raudaschl, P. Zaffino, M. F. Spadea, G. C. Sharp, and R. Schubert, "Deep neural networks for fast segmentation of 3D medical images," in *International Conference on Medical Image Computing and Computer-Assisted Intervention*, 2016, pp. 158-165.

[6] T. Doshi, J. Soraghan, L. Petropoulakis, G. Di Caterina, D. Grose, K. MacKenzie*, et al.*, "Automatic pharynx and larynx cancer segmentation framework (PLCSF) on contrast enhanced MR images," *Biomedical Signal Processing and Control,* vol. 33, pp. 178-188, 2017.

[7] Y. Zhang, B. J. Matuszewski, L.-K. Shark, and C. J. Moore, "Medical image segmentation using new hybrid level-set method," in *2008 fifth international conference biomedical visualization: information visualization in medical and biomedical informatics*, 2008, pp. 71-76.

[8] B. Zhao, J. Soraghan, D. Grose, T. Doshi, and G. Di-Caterina, "Automatic 3D Detection and Segmentation of Head and Neck Cancer from MRI Data," in *2018 7th European Workshop on Visual Information Processing (EUVIP)*, 2018, pp. 1-6.

[9] G. Litjens, T. Kooi, B. E. Bejnordi, A. A. A. Setio, F. Ciompi, M. Ghafoorian*, et al.*, "A survey on deep learning in medical image analysis," *Medical image analysis,* vol. 42, pp. 60-88, 2017.

[10] Y. LeCun, L. Bottou, Y. Bengio, and P. Haffner, "Gradient-based learning applied to document recognition," *Proceedings of the IEEE,* vol. 86, pp. 2278-2324, 1998.

[11] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," in *Advances in neural information processing systems*, 2012, pp. 1097-1105.

[12] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," *arXiv preprint arXiv:1409.1556,* 2014.

[13] J. Long, E. Shelhamer, and T. Darrell, "Fully convolutional networks for semantic segmentation," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2015, pp. 3431-3440.

[14] O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," in *International Conference on Medical image computing and computer-assisted intervention*, 2015, pp. 234-241.

[15] Ö. Çiçek, A. Abdulkadir, S. S. Lienkamp, T. Brox, and O. Ronneberger, "3D U-Net: learning dense volumetric segmentation from sparse annotation," in *International conference on medical image computing and computer-assisted intervention*, 2016, pp. 424-432.

[16] H. Dong, G. Yang, F. Liu, Y. Mo, and Y. Guo, "Automatic brain tumor detection and segmentation using U-Net based fully convolutional networks," in *annual conference on medical image understanding and analysis*, 2017, pp. 506-517.

[17] F. Yu and V. Koltun, "Multi-scale context aggregation by dilated convolutions," *arXiv preprint arXiv:1511.07122,* 2015.

[18] F. Milletari, N. Navab, and S.-A. Ahmadi, "V-net: Fully convolutional neural networks for volumetric medical image segmentation," in *2016 Fourth International Conference on 3D Vision (3DV)*, 2016, pp. 565-571.

[19] L. R. Dice, "Measures of the amount of ecologic association between species," *Ecology,* vol. 26, pp. 297-302, 1945.

[20] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," *arXiv preprint arXiv:1412.6980,* 2014.

[21] S. Ioffe and C. Szegedy, "Batch normalization: Accelerating deep network training by reducing internal covariate shift," *arXiv preprint arXiv:1502.03167,* 2015.

.