Deep Learning-based Automated Delineation of Head and Neck Malignant Lesions from PET Images

Hossein Arabi, Isaac Shiri, Elnaz Jenabi, Minerva Becker, and Habib Zaidi, *Fellow, IEEE*

2020 IEEE Nuclear Science Symposium and Medical Imaging Conference (NSS/MIC) | 978-1-7281-7693-2/20/$31.00 ©2020 IEEE | DOI: 10.1109/NSS/MIC42677.2020.9507977

***Abstract*– Accurate delineation of the gross tumor volume (GTV) is critical for treatment planning in radiation oncology. This task is very challenging owing to the irregular and diverse shapes of malignant lesions. Manual delineation of the GTVs on PET images is not only time-consuming but also suffers from inter- and intra-observer variability. In this work, we developed deep learning-based approaches for automated GTV delineation on PET images of head and neck cancer patients. To this end, V-Net, a fully convolutional neural network for volumetric medical image segmentation, and HighResNet, a 20-layer residual convolutional neural network, were adopted. ^{18}F-FDG-PET/CT images of 510 patients presenting with head and neck cancer on which manually defined (reference) GTVs were utilized for training, evaluation and testing of these algorithms. The input of these networks (in both training or evaluation phases) were 12×12×12 cm sub-volumes of PET images containing the whole volume of the tumors and the neighboring background radiotracer uptake. These networks were trained to generate a binary mask representing the GTV on the input PET sub-volume. Standard segmentation metrics, including Dice similarity and precision were used for performance assessment of these algorithms. HighResNet achieved automated GTV delineation with a Dice index of 0.87±0.04 compared to 0.86±0.06 achieved by V-Net. Despite the close performance of these two approaches, HighResNet exhibited less variability among different subjects as reflected in the smaller standard deviation and significantly higher precision index (0.87±0.07 versus 0.80±0.10). Deep learning techniques, in particular HighResNet algorithm, exhibited promising performance for automated GTV delineation on head and neck PET images. Incorporation of anatomical/structural information, particularly MRI, may result in higher segmentation accuracy or less variability among the different subjects.**

Index Terms*— Head and Neck Cancer, Segmentation, PET, Deep Learning

I. Introduction

HEAD and neck cancer is among the most prevalent types of cancer worldwide. High precision external radiation therapy (RT) is considered an effective strategy for treatment of this type of cancer. A key step to conduct an effective image-guided RT planning is accurate delineation of the gross tumor volume (GTV).

Accurate and robust GTV delineation allows for maximum therapeutic dose delivery to the target volume and sparing healthy tissues from unwanted toxicities. This issue is of special importance in head and neck tumors owing to the proximity to critical anatomical structures. In clinical practice, the GTV delineation is often performed by radiation oncologists via manual or semi-manual techniques [1]. This procedure is not only time-consuming but also prone to inter- and intra-observer variability. A number of studies have shown that multi-observer delineation of the GTV vary significantly due to the operator's knowledge, experience and workload [2]. In this light, an accurate, robust and reproducible GTV delineation method is crucial for effective RT planning for head and neck tumors.

Automated organ and tumor delineation, GTV segmentation of head and neck tumors is challenging due to their irregular shapes, proximity to critical structures and the presence of healthy tissues presenting with similar appearance/signals [3-6]. Deep learning-based approaches have exhibited promising performance in medical image analysis, including image artifact reduction [7, 8], image enhancement [9-11], quantification [12] and segmentation [13, 14]. In this work, we develop deep learning-based approaches for automated GTV delineation from PET images of the head and neck region. To this end, V-Net, a fully convolutional neural network for volumetric medical image segmentation, and HighResNet, a 20-layer residual convolutional neural network, algorithms were adopted. The aim of this study is to examine the level of segmentation accuracy, overall performance and variability across different subjects achievable by deep learning approaches.

II. MATERIALS AND METHODS

A. Description of PET/CT datasets

The datasets consisted of ^{18}F-FDG PET/CT images of 510 patients presenting with head and neck cancer obtained from the cancer imaging archive (TCIA) [15]. The corresponding GTVs were manually delineated on PET images by an experienced nuclear medicine physician and saved in binary format. To feed PET images into the deep learning algorithms, 12×12×12 cm sub-volumes of PET images containing the whole volume of the tumors and neighboring background radiotracer uptake were extracted from the original PET images to create a uniform dataset in terms of matrix and voxel size. The binary mask, representing GTVs, were processed accordingly.

Manuscript was submitted December 20, 2020. This work was supported by the Swiss National Science Foundation under grants SNSF 320030_176052 and 320030_173091.

H. Arabi, I. Shiri, M. Becker and H. Zaidi are with the Division of Nuclear Medicine & Molecular Imaging, Geneva University Hospital, Geneva, Switzerland (e-mail: hossein.arabi@unige.ch, Isaac.shirilord@unige.ch, Minerva.Becker@unige.ch, habib.zaidi@hcuge.ch).

E. Jenabi is with Research Centre for Nuclear Medicine, Shariati Hospital, Tehran University of Medical Sciences, Tehran, Iran (e-mail: Jenabi.elnaz@gmail.com)

Habib Zaidi is with Geneva University Neurocenter, Geneva University, Geneva, Switzerland (e-mail: habib.Zaidi@hcuge.ch).

978-1-7281-7693-2/20/$31.00 ©2020 IEEE

B. Deep learning approaches

Two state-of-the-art deep learning approaches (V-Net and HighResNet) with promising performance in segmentation of the medical images were implemented in this study [16, 17]. V-Net is a fully convolutional neural network for volumetric medical image segmentation consisting of 10 stages with different receptive fields varying from 5 ×5 ×5 to 551×551×551. HighResNet is a 20-layer residual convolutional neural network equipped with dilated convolutional operations. These networks were trained in a 10-fold cross validation scheme using Dice index as the loss function and batch size of 1. The cropped PET images scaled in SUV units were considered as input to these algorithms to generate the corresponding binary masks of GTVs as output.

C. Clinical evaluation

Manual segmentation of the GTVs was considered as reference based on which the performance of the V-Net and HiResNet models was evaluated. To this end, the Dice index (Eq. 1), Jaccard similarity coefficient (Eq. 2), Precision (Eq. 3), and Matthew correlation coefficient (MCC) (Eq. 4) were calculated across the 510 subjects.

$$DSC = \frac{2|A \cap M|}{|A|+|M|} \quad \text{(Eq. 1)}$$

$$Jaccard = \frac{|A \cap M|}{|A \cup M|} \quad \text{(Eq. 2)}$$

$$Precision = \frac{True\ Positive}{True\ Positive + False\ positive} \quad \text{(Eq. 3)}$$

$$MCC = \frac{TP \times TN - FP \times FN}{\sqrt{(TP+FP)\times(TP+FN)\times(TN+FP)\times(TN+FN)}} \quad \text{(Eq. 4)}$$

III. Results and Discussion

Fig. 1 depicts representative examples GTV delineation on PET images using V-Net and HighResNet algorithms. Three-dimensional rendered volumes of tumors are also presented for visual inspection.

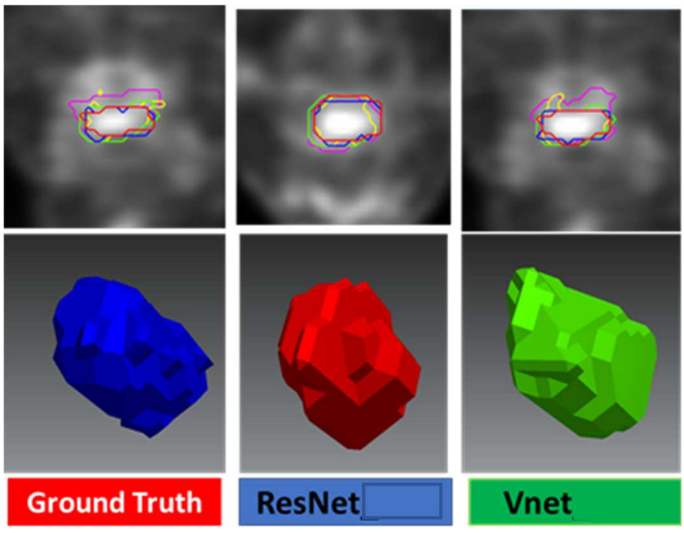

Fig. 1. Representative GTV delineation examples on PET images using V-Net and HighResNet algorithms.

Fig 2 presents box plots of segmentation metrics, including Dice, Jaccard, Percision and MCC, calculated for the HighResNet and V-Net algorithms across 510 subjects. Regarding the Dice index, both deep learning algorithms exhibited promising performance. However, HighResNet performed slightly better with a Dice of 0.87±0.04 compared to 0.86±0.06 achieved by V-Net. Moreover, HighResNet exhibited less variability and more robustness across the different subjects, which is reflected by a lower standard deviation.

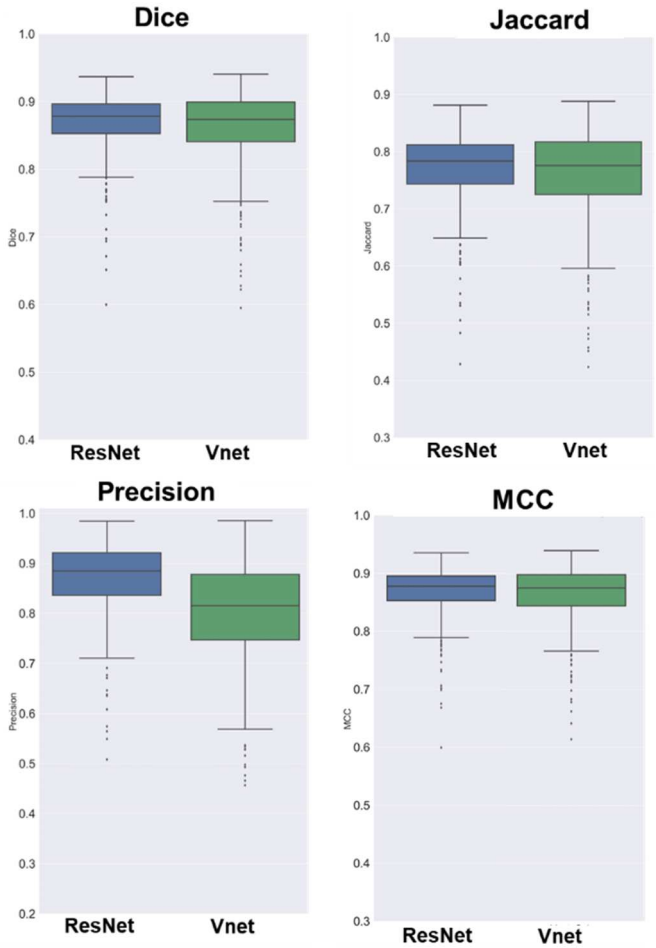

Fig 2. Boxplot of different parameters for ResNet and Vnet

TABLE I. summary of pet segmentation metrics calculated across 510 subjects for ResNet and Vnet models.

Quantitative Parameters	ResNet	Vnet
Precision	0.87±0.07	0.80±0.10
MCC	0.87±0.04	0.86±0.05
Dice	0.87±0.04	0.86±0.06
Jaccard	0.77±0.06	0.76±0.08

Table 1 summarizes the statistical summary of the segmentation metrics for HighResNet and V-Net algorithms. Regarding the Precision metric, the HighResNet algorithm showed superior performance over the V-Net model. HighResNet algorithm resulted in a Precision of 0.87±0.07, while V-Net model led to a Precision of 0.80±0.10. Not only the overall Precision across the 510 subjects is higher in HighResNet model, a lower standard deviation demonstrated the robustness of this model for automated GTV delineation.

IV. CONCLUSION

Deep learning techniques, in particular HighResNet algorithm, exhibited promising performance for automated GTV delineation on head and neck PET images. HighResNet algorithm not only resulted in slightly better quantification of lesion volumes but also showed less variability in performance across the different subjects. Incorporation of anatomical/structural information, particularly MRI, may result in higher segmentation accuracy or less variability among the different subjects.

REFERENCES

[1] M. Hatt *et al.*, "Classification and evaluation strategies of auto-segmentation approaches for PET: Report of AAPM Task Group No. 211.," *Med Phys,* vol. 44, no. 6, pp. e1-e42, Jan 24 2017, doi: 10.1002/mp.12124.

[2] P. M. Harari, S. Song, and W. A. Tomé, "Emphasizing conformal avoidance versus target definition for IMRT planning in head-and-neck cancer," (in eng), *International journal of radiation oncology, biology, physics,* vol. 77, no. 3, pp. 950-8, Jul 1 2010, doi: 10.1016/j.ijrobp.2009.09.062.

[3] Z. Guo, N. Guo, K. Gong, S. Zhong, and Q. Li, "Gross tumor volume segmentation for head and neck cancer radiotherapy using deep dense multi-modality network," (in eng), *Phys Med Biol,* vol. 64, no. 20, p. 205015, Oct 16 2019, doi: 10.1088/1361-6560/ab440d.

[4] H. Arabi and H. Zaidi, "Applications of artificial intelligence and deep learning in molecular imaging and radiotherapy," *European Journal of Hybrid Imaging,* vol. 4, no. 1, p. 17, 2020/09/23 2020, doi: 10.1186/s41824-020-00086-8.

[5] R. Mohammadi, I. Shokatian, M. Salehi, H. Arabi, I. Shiri, and H. Zaidi, "Deep learning-based auto-segmentation of organs at risk in high-dose rate brachytherapy of cervical cancer," (in eng), *Radiother Oncol,* vol. 159, pp. 231-240, Apr 6 2021, doi: 10.1016/j.radonc.2021.03.030.

[6] H. Arabi, A. AkhavanAllaf, A. Sanaat, I. Shiri, and H. Zaidi, "The promise of artificial intelligence and deep learning in PET and SPECT imaging," (in eng), *Phys Med,* vol. 83, pp. 122-137, Mar 22 2021, doi: 10.1016/j.ejmp.2021.03.008.

[7] I. Shiri, P. Sheikhzadeh, and M. R. Ay, "Deep-fill: Deep learning based sinogram domain gap filling in positron emission tomography," *arXiv preprint arXiv:1906.07168,* 2019.

[8] I. Shiri *et al.*, "Deep-JASC: joint attenuation and scatter correction in whole-body (18)F-FDG PET using a deep residual network," (in eng), *Eur J Nucl Med Mol Imaging,* vol. 47, no. 11, pp. 2533-2548, Oct 2020, doi: 10.1007/s00259-020-04852-5.

[9] I. Shiri *et al.*, "Ultra-low-dose chest CT imaging of COVID-19 patients using a deep residual neural network," (in eng), *Eur Radiol,* pp. 1-12, Sep 3 2020, doi: 10.1007/s00330-020-07225-6.

[10] I. Shiri *et al.*, "Standard SPECT myocardial perfusion estimation from half-time acquisitions using deep convolutional residual neural networks," (in eng), *J Nucl Cardiol,* Apr 28 2020, doi: 10.1007/s12350-020-02119-y.

[11] A. Sanaat, I. Shiri, H. Arabi, I. Mainta, R. Nkoulou, and H. Zaidi, "Deep learning-assisted ultra-fast/low-dose whole-body PET/CT imaging," (in eng), *Eur J Nucl Med Mol Imaging,* Jan 25 2021, doi: 10.1007/s00259-020-05167-1.

[12] A. Akhavanallaf, I. Shiri, H. Arabi, and H. Zaidi, "Whole-body voxel-based internal dosimetry using deep learning," (in eng), *Eur J Nucl Med Mol Imaging,* Sep 1 2020, doi: 10.1007/s00259-020-05013-4.

[13] S. Moradi *et al.*, "MFP-Unet: A novel deep learning based approach for left ventricle segmentation in echocardiography," *Physica Medica,* vol. 67, pp. 58-69, 2019/11/01/ 2019, doi: https://doi.org/10.1016/j.ejmp.2019.10.001.

[14] I. Shiri *et al.*, "COLI-NET: Fully Automated COVID-19 Lung and Infection Pneumonia Lesion Detection and Segmentation from Chest CT Images," *medRxiv,* p. 2021.04.08.21255163, 2021, doi: 10.1101/2021.04.08.21255163.

[15] K. Clark *et al.*, "The Cancer Imaging Archive (TCIA): maintaining and operating a public information repository," *Journal of digital imaging,* vol. 26, no. 6, pp. 1045-1057, 2013.

[16] 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: IEEE, pp. 565-571.

[17] W. Li, G. Wang, L. Fidon, S. Ourselin, M. J. Cardoso, and T. Vercauteren, "On the compactness, efficiency, and representation of 3D convolutional networks: brain parcellation as a pretext task," in *International conference on information processing in medical imaging*, 2017: Springer, pp. 348-360.