

# Cardio-pulmonary Substructure Segmentation of CT images using Convolutional Neural Networks for Clinical Outcome Analysis

Rabia Haq<sup>1</sup>  
Alexandra Hotca<sup>1</sup>  
Aditya Apte<sup>1</sup>  
Andreas Rimner<sup>2</sup>  
Joseph O. Deasy<sup>1</sup>  
Maria Thor<sup>1</sup>

HAQR@MSKCC.ORG

<sup>1</sup> *Department of Medical Physics, Memorial Sloan Kettering Cancer Center, New York, US*

<sup>2</sup> *Department of Radiation Oncology, Memorial Sloan Kettering Cancer Center, New York, US*

**Editors:** Under Review for MIDL 2019

**Keywords:** Semantic segmentation, Convolutional Neural Networks, cardio-pulmonary, clinical outcomes

## 1. Introduction

Various studies have shown that doses to some cardio-vascular substructures may be critical factors in the observed heart toxicity and early mortality following radiotherapy (RT) for nonsmall cell lung cancer (NSCLC) patients (Vivekanandan et al., 2017; McWilliam et al., 2017; RT et al., 2017; Thor et al., 2018). This may be attributed to irradiation of particular constituents of the cardio-pulmonary system [2-5]. Currently, segmentation of cardio-pulmonary organs other than the whole heart and lung has been overlooked, and only these two organs are routinely defined as part of the treatment planning process. RT planning requires robust and accurate segmentation of organs-at-risk in order to maximize radiation to the disease location and spare the normal tissue as much as possible. The introduction of a new set of organs puts requirements on both segmentation accuracy and segmentation time that would result in an overhead of several hours of manual segmentation and contour refinement in the clinic.

To facilitate this, we built and validated a multi-label Deep Learning Segmentation (DLS) framework for accurate auto-segmentation of cardio-pulmonary substructures. The DLS framework utilized a deep convolutional neural network architecture to segment 12 cardio-pulmonary substructures (Feng et al., 2010) from Computed Tomography (CT) scans of 217 patients previously treated with thoracic RT. The segmented substructures are: Heart, Pericardium, Atria, Ventricles, Descending Aorta (DA), Left Atrium (LA), Right Atrium (RA), Left Ventricle (LA), Right Ventricle (RV), Inferior Vena Cava (IVC), Superior Vena Cava (SVC) and Pulmonary Artery (PA). We evaluate our framework using a hold-out dataset of 24 CT scans by calculating volumetric-based as well as dose-volume histogram (DVH) based validation metrics. The proposed model reduces substructure seg-

mentation time for a new patient from about one hour of manual segmentation to approximately 10 seconds. We demonstrate that the model is robust against variability in image quality characteristics, including the presence/absence of contrast. Resulting segmentation accuracy was judged adequate for extracting dose-volume histogram information for patient outcomes analyses following RT, with no statistical difference discovered between auto-generated and expert contour evaluation metrics.

## 2. Methods

Experimental data consisted of computerized tomography (CT) scans of 241 patients obtained from our institutional clinic. This data consisted of contrast as well as non-contrast enhanced images of varying imaging quality and resolution across different scanners. Manual expert segmentation for 12 organs-at-risk cardio-pulmonary structures was considered ground truth and used for model training, testing and validation.

Our approach leverages the deep neural network architecture of (Chen et al., 2018). Convolutional neural networks (CNNs) and encoder-decoder neural networks have been successfully employed for medical image segmentation tasks (Isensee et al., 2018), (Oktay et al., 2018), (Jin et al., 2018), (Oktay et al., 2018). The Deeplab encoder-decoder network architecture with atrous separable convolutions consists of spatial pyramid pooling that encodes multi-scale contextual information to capture spatial anatomical information of contiguous structures. Dense feature maps extracted in the last encoder network path consist of detailed semantic information. The decoder network is able to robustly recover structure boundaries through bilinear upsampling at a factor of 4 while applying atrous convolutions to reduce features before semantic labeling. We trained the network using ResNet-101(He et al., 2016) as the encoder network backbone with learning rate = 0.01 using policy learning rate scheduler (Liu et al., 2015), crop size= $513 \times 513$ , batch size = 8, loss = cross-entropy, output stride = 16 for 50 epochs for dense label prediction. Our approach has been implemented using the Pytorch DL framework.

We quantitatively evaluated the auto-generated segmentations by comparing the DSC Score and 95th Percentile Hausdorff Distance (HD95 (mm)) of 24 patients against expert clinical segmentations. Additionally, we calculated the difference in RT DVH metrics between auto-generated and expert contours. The wilcoxon rank-sum test was performed to determine any statistical difference between the metrics.

## 3. Experiments and Results

Figure 1 displays the DSC Score results for 24 hold-out validation CT images. Our achieved DSC accuracies are comparable to the state-of-the-art multi-atlas (Luo et al., 2019) and deep learning methods (Dormer and et al., 2018) for segmenting cardio-pulmonary substructures from CT images. The highest segmentation accuracy was observed for the heart (median DSC = 0.96, median HD95= 3.48 mm), while the remaining structures achieving median accuracy ( $0.81 \leq \text{DSC} \leq 0.94$ ) and ( $6 \text{ mm} \leq \text{HD95} \leq 3 \text{ mm}$ ), with highest HD95 surface distance accuracy observed for DA.

Table 1 displays the percentage difference between the DVH metrics and their associated P-values for six substructure contours that are found to be critical for determining heart

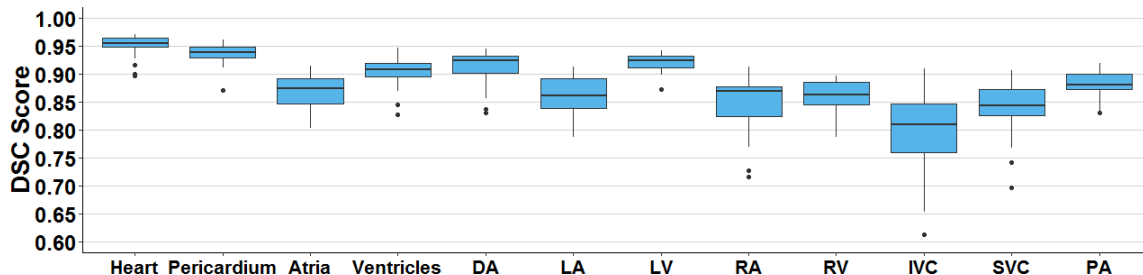


Figure 1: Median Dice Similarity Coefficient (DSC) Score results of 24 thoracic RT CT images comparing auto-generated DLS contours against manually segmented expert contours for 12 cardio-pulmonary sub-structures.

toxicity. These metrics were calculated using the RT treatment plan generated by physicists for each patient. None of the structure DVH metric differences were found to be statistically significant, with all P-values > 0.05. This indicates that the auto-generated segmentations are clinically acceptable for outcome analysis and RT treatment planning in the clinic.

Structure (metric)	DLS to Expert Contour Difference (%)	P-value
Atria D45% (Gy)	0.00 (-5.19 8.86)	0.89
Heart V2 (%)	0.00 (0.00 0.01)	0.93
Heart V50 (%)	0.00 (-0.05 0.10)	0.83
Left Atrium Dmax (Gy)	0.00 (-0.28 0.31)	0.90
Left Atrium V63-V69 (%)	-0.06 (-0.25 0.75)	0.92
Pericardium MOH55% (Gy)	1.76 (0.24 4.10)	0.69
Superior Vena Cava D90% (Gy)	0.00 (-1.51 0.09)	0.83
Ventricles MOH5% (Gy)	-0.14 (-4.53 1.66)	0.98

Table 1: Comparing Dose Volume Histogram (DVH) metrics of auto-generated DLS contours against Expert contours for 6 structures. Median and inter-quartile range of percentage differences between the two contours is presented, which is calculated as:  $(DLS\_Volume - Expert\_Volume / Expert\_Volume) \times 100$ .

#### 4. Conclusion

We propose a model for auto-segmentation of cardio-pulmonary substructures from contrast and non-contrast enhanced CT images. We validated our approach by quantitatively comparing resulting contours against expert delineation, and further demonstrated no statistical difference when used for dose-volume histogram calculations. Resulting segmentations can effectively be utilized to study the effect of heart toxicity and clinical outcomes, as well as used as input to RT treatment planning. We have applied our approach to auto-segment additional 283 treatment planning CT scans to study heart toxicity outcomes in non-advanced lung patients. The developed cardio-pulmonary segmentation models are being integrated into deep learning tools within the open-source CERR (Deasy et al., 2003) platform.

## Acknowledgments

This research is partially supported by NCI R01 CA198121.

## References

- Liang-Chieh Chen, Yukun Zhu, George Papandreou, Florian Schroff, and Hartwig Adam. Encoder-decoder with atrous separable convolution for semantic image segmentation. In *ECCV*, 2018.
- JO Deasy, AI Blanco, and VH Clark. Cerr: a computational environment for radiotherapy research. *Med Phys.*, 30(5):979–85, 2003.
- JD Dormer and L Ma et al. Heart chamber segmentation from ct using convolutional neural networks. In *Proc SPIE Int Soc Opt Eng.*, 2018.
- M Feng, JM Moran, T Koelling, and et al. Development and validation of a heart atlas to study cardiac exposure to radiation following treatment for breast cancer. *Int J Radiat Oncol Biol Phys*, 79(1):1018, 2010.
- K. He, X. Zhang, S. Ren, and J. Sun. Deep residual learning for image recognition. In *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 770–778, June 2016. doi: 10.1109/CVPR.2016.90.
- Fabian Isensee, Jens Petersen, André Klein, David Zimmerer, Paul F. Jaeger, Simon Kohl, Jakob Wasserthal, Gregor Koehler, Tobias Norajitra, Sebastian J. Wirkert, and Klaus H. Maier-Hein. nnu-net: Self-adapting framework for u-net-based medical image segmentation. *CoRR*, abs/1809.10486, 2018. URL <http://arxiv.org/abs/1809.10486>.
- Qiangguo Jin, Zhao-Peng Meng, Changming Sun, Leyi Wei, and Ran Su. Ra-unet: A hybrid deep attention-aware network to extract liver and tumor in CT scans. *CoRR*, abs/1811.01328, 2018. URL <http://arxiv.org/abs/1811.01328>.
- Wei Liu, Andrew Rabinovich, and Alexander C. Berg. Parsenet: Looking wider to see better. *CoRR*, abs/1506.04579, 2015. URL <http://arxiv.org/abs/1506.04579>.
- Yangkun Luo, Yujin Xu, Zhongxing Liao, Daniel Gomez, Jingqian Wang, Wei Jiang, Rongrong Zhou, Ryan Williamson, Laurence E. Court, and Jinzhong Yang. Automatic segmentation of cardiac substructures from noncontrast ct images: accurate enough for dosimetric analysis? *Acta Oncologica*, 58(1):81–87, 2019.
- A McWilliam, , J Kennedy, and C Hodgson et al. Radiation dose to heart base linked with poorer survival in lung cancer patients. *Eur J Cancer*, 85:10613, 2017.
- O. Oktay, E. Ferrante, K. Kamnitsas, M. Heinrich, W. Bai, J. Caballero, S. A. Cook, A. de Marvaio, T. Dawes, D. P. ORegan, B. Kainz, B. Glocker, and D. Rueckert. Anatomically constrained neural networks (acnns): Application to cardiac image enhancement and segmentation. *IEEE Transactions on Medical Imaging*, 37(2):384–395, Feb 2018.

Ozan Oktay, Jo Schlemper, Loïc Le Folgoc, Matthew C. H. Lee, Mattias P. Heinrich, Kazunari Misawa, Kensaku Mori, Steven G. McDonagh, Nils Y. Hammerla, Bernhard Kainz, Ben Glocker, and Daniel Rueckert. Attention u-net: Learning where to look for the pancreas. *CoRR*, abs/1804.03999, 2018. URL <http://arxiv.org/abs/1804.03999>.

RT Dess RT, Y Sun, and MM Matuszak et al. Cardiac events after radiation therapy: combined analysis of prospective multicenter trials for locally advanced non-small-cell lung cancer. 35:1395402, 2017.

M Thor, JO Deasy, and C Hu et al. The role of heart-related dose-volume metrics on overall survival in the rtog 0617 clinical trial. *Int J Radiat Oncol Biol Phys*, page 102:S96, 2018.

S Vivekanandan, DB Landau, N Counsell, DR Warren, A Khwanda, and SD Rosen et al. The impact of cardiac radiation dosimetry on survival after radiation therapy for non-small cell lung cancer. *Int J Radiat Oncol*, 99:5160, 2017.