Title: Deep learning-based auto-segmentation of swallowing and chewing structures

Authors: Aditi Iyer[1], Maria Thor[1], Rabia Haq[1], Joseph O. Deasy[1], Aditya P. Apte[1]

[1]*Department of Medical Physics, Memorial Sloan Kettering Cancer Center, New York, USA*

Abstract

Purpose

Delineating the swallowing and chewing structures in Head and Neck (H&N) CT scans is necessary for radiotherapy treatment (RT) planning to reduce the incidence of radiation-induced dysphagia, trismus, and speech dysfunction. Automating this process would decrease the manual input required and yield reproducible segmentations, but generating accurate segmentations is challenging due to the complex morphology of swallowing and chewing structures and limited soft tissue contrast in CT images.

Methods

We trained deep learning models using 194 H&N CT scans from our institution to segment the masseters (left and right), medial pterygoids (left and right), larynx, and pharyngeal constrictor muscle using DeepLabV3+ with the resnet-101 backbone. Models were trained in a sequential manner to guide the localization of each structure group based on prior segmentations. Additionally, an ensemble of models was developed using contextual information from three different views (axial, coronal, and sagittal), for robustness to occasional failures of the individual models. Output probability maps were averaged, and voxels were assigned labels corresponding to the class with the highest combined probability.

Results

The median dice similarity coefficients (DSC) computed on a hold-out set of 24 CT scans were 0.87±0.02 for the masseters, 0.80±0.03 for the medial pterygoids, 0.81±0.04 for the larynx, and 0.69±0.07for the constrictor muscle. The corresponding 95th percentile Hausdorff distances were

0.32±0.08cm (masseters), 0.42±0.2cm (medial pterygoids), 0.53±0.3cm (larynx), and 0.36±0.15cm (constrictor muscle). Dose-volume histogram (DVH) metrics previously found to correlate with each toxicity were extracted from manual and auto-generated contours and compared between the two sets of contours to assess clinical utility. Differences in DVH metrics were not found to be statistically significant ($p>0.05$) for any of the structures. Further, inter-observer variability in contouring was studied in 10 CT scans. Automated segmentations were found to agree better with each of the observers as compared to inter-observer agreement, measured in terms of DSC.

Conclusions

We developed deep learning-based auto-segmentation models for swallowing and chewing structures in CT. The resulting segmentations can be included in treatment planning to limit complications following RT for H&N cancer. The segmentation models developed in this work are distributed for research use through the open-source platform CERR, accessible at https://github.com/cerr/CERR.

1. INTRODUCTION

Delineating organs at risk (OAR) is central in radiotherapy (RT) treatment planning to limit normal tissue complications following RT. Manual delineation is time-consuming, subjective, and prone to errors due to factors such as organ complexity and level of experience [1]. Therefore, there has been a great deal of interest in automating this process to generate accurate and reproducible segmentations in a time-efficient manner.

In head and neck (H&N) cancer treatment, isolating the larynx and pharyngeal constrictor muscle is of particular interest to limit speech dysfunction and dysphagia, respectively, following RT [2] [3]. The masseters and medial pterygoids have also been identified as critical structures in limiting radiation-induced trismus [4]. However, delineating these structures is challenging due to their complex morphology and the low soft tissue contrast in CT images.

Few semi-automatic and automatic methods have been previously developed to segment OARs in the H&N. Conventional multi-atlas-based auto-segmentation (MABAS) methods involve propagating and combining manually-segmented OARs from a curated library of CT scans through image registration as in [5] and [6]. Other approaches offer strategies to further refine MABAS using organ-specific intensity [7] and texture [8] features or shape representation models [9] [10]. However, MABAS is sensitive to inter-subject anatomical variations as well as image artifacts, and image registration is computationally intensive, requiring several minutes even with highly efficient implementations [11].

Convolutional neural networks (CNNs) have recently been applied successfully to various medical image segmentation tasks. Ibragimov et al. [12] trained 13 CNNs, applied in sliding-window fashion to segment H&N OARs including the larynx and pharynx, which were further refined using Markov Random Field (MRF)-based post-processing. In [13], Ward van Rooij et al. employed the popular 3D U-net [14] architecture to segment H&N OARs including the pharyngeal constrictor muscle. Zhu et all. [11] extended the U-Net model by incorporating squeeze-and excitation (SE) residual blocks and a modified loss function to improve segmentation of smaller structures such as the chiasm and optic nerves. In [15], Men et al. segmented nasopharyngeal tumor volumes in H&N CT using a modified version of the VGG-16 [16] architecture, replacing fully-connected layers with fully-convolutional layers and introducing improved decoder networks to rebuild high-resolution feature maps. Tong et al [17] trained a fully convolutional neural net (FCNN), incorporating prior information by training a shape representation model to regularize shape characteristics of 9 H&N OARs. The FocusNet [18] developed by Gao et al. utilizes multiple CNNs to segment H&N OARs including the larynx, first segmenting large structures, then segmenting smaller structures with specifically designed sub-networks.

In this work, we present a fully automatic method to segment swallowing and chewing structures and examine its suitability for clinical use. To the best of our knowledge, this is the first [35] deep learning-based method for segmenting the chewing structures in CT images. We propose a novel framework in which DeepLabV3+ segmentation models are trained sequentially to guide the localization of each structure group based on previously-segmented structures. Model ensembles are created using three orthogonal views (axial, sagittal, and coronal) in 2.5D and shown to improve segmentation accuracy of H&N OARs compared to models trained on a single-view.

2. MATERIALS AND METHODS

2.1 Dataset

CT scans of 243 H&N cancer patients from our institution were retrospectively collected under IRB 16-142 and accessed under IRB 16-1488 to develop the auto-segmentation methods in this work. The dataset was randomly partitioned into training (80%), validation (10%), and testing (10%) sets. The validation set (10%) was used to estimate model performance while tuning hyperparameters, and the testing set (10%) was used for unbiased evaluation of the final model. Representative CT images showing considerable variation in head pose, shape, and appearance around the structures of interest, including slices with dental artifacts due to dental implants, were included in the dataset. Additional data characteristics are listed in Table 1. Manual segmentations of the masseters (left, right), medial pterygoids (left, right), larynx, and constrictor muscle, generated using MSKCC's in-house treatment planning system, served as the reference standard. Reference contours of the masseters and medial pterygoids were available in 60% of the scans, and the constrictor muscle in 97% of the scans (larynx was provided in all scans).

Table 1. Summary of data characteristics for axial images

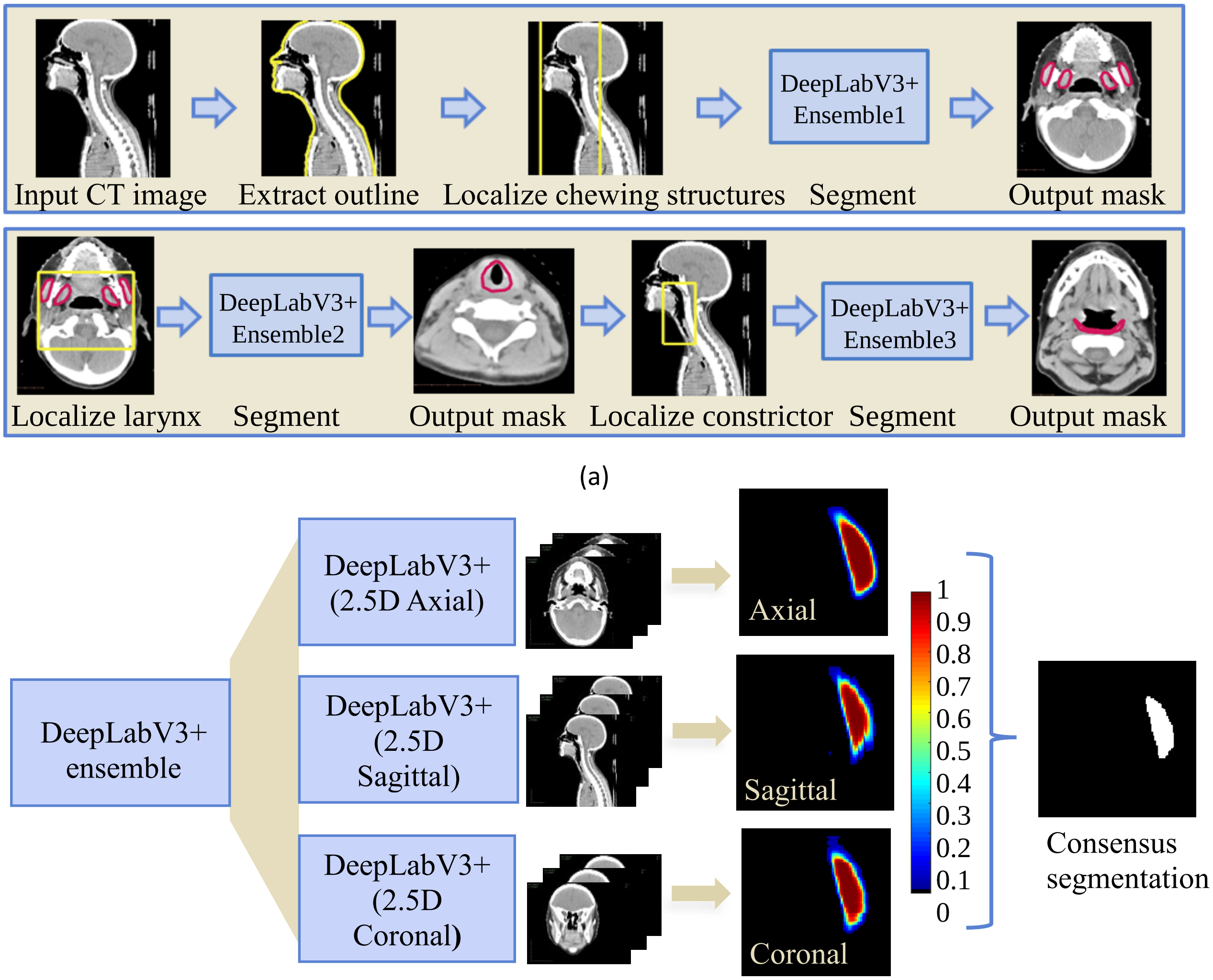
Attribute	**Median**	**Min**	**Max**	**5th percentile**	**95th percentile**
Number of slices	80	48	279	65	103
In-plane spacing (mm)	0.78	0.51	1.37	0.61	1.17
Slice thickness (mm)	3.00	1.25	6.00	2.50	3.27

2.2 CNN segmentation model

The DeepLabV3+ CNN [19] was selected due to its impressive performance on the PASCAL VOC 2012 and Cityscapes datasets. Moreover, it has previously been applied successfully to medical image segmentation tasks, e.g., by Elguindi et al. [26] to segment the prostate in MR images and by Haq et al. [27] to segment cardio-pulmonary substructures in CT images. This architecture offers the advantage of an encoder network that can capture contextual information at different scales using multiple dilated convolution layers applied in parallel at different rates and a decoder network capable of effectively recovering object boundaries.

Here, training was performed using ResNet-101 [20] as the backbone of the encoder network. A publicly-distributed implementation [21] of DeepLabV3+ using the Pytorch [33] framework was utilized, and a soft-max layer was appended to obtain voxel-wise probabilities. On training a single multi-class model to segment all the structures of interest, it was observed that the imbalance in class labels due to large differences in OAR sizes led to poor performance on smaller structures. Augmenting the loss function based on class frequencies did not produce significant improvement, as previously observed in [18]. Consequently, three separate models were developed- one to segment the chewing structures and one each for the larynx and the constrictor muscle. A sequential segmentation strategy (figure 1) was utilized in which each segmented OAR was used to constrain the location of subsequently segmented OARs. Additionally, an ensemble of three models was developed per OAR group using 2D axial, sagittal, and coronal slices. As compared to 3D convolutional neural networks (CNNs) which are highly memory-intensive, training in 2D enabled us to employ a more complex CNN while still providing contextual information and increased redundancy from the three different orientations.

A high-performance cluster with four NVIDIA GeForce GTX 1080Ti GPUs with 11GB of memory each was used in training. Batch-normalization was applied using mini-batches of 8 images, resized to 320 x 320 voxels. Input images were standardized by scaling the intensities to [0,1] and normalizing to zero-mean and unit-variance. Data augmentation was performed through image scaling, cropping, and rotation. Input channels were populated using three consecutive slices (axial, sagittal, and coronal, respectively, for the three models), and the cross-entropy loss was employed in training. Data preprocessing and export to HDF5 [28] format was performed using the Computational Environment for Radiological Research (CERR) [23]. Details of the individual models are presented in the following sections, and Table 2 summarizes the hyperparameters used.



(a)

(b)

Figure 1. (a) Sequential framework for segmenting chewing and swallowing structures using deep learning, in which each segmented OAR group is used to improve localization of subsequently segmented OARs. (b) Example showing consensus segmentation of left masseter using ensemble model trained on 3 orthogonal views (axial, sagittal, and coronal).

2.3 Chewing structures

A multi-label model was trained to segment the left and right masseters and medial pterygoids. H&N CT scans were automatically cropped around the patient's outline prior to training. The couch was first detected using the Hough transform and masked out. This was followed by intensity-based thresholding and morphological post-processing to extract the patient's outline. A bounding box was generated around the outline and its posterior extent was limited by 25%. 2D axial, sagittal, and coronal images and corresponding masks of the chewing structures were extracted within these extents for training.

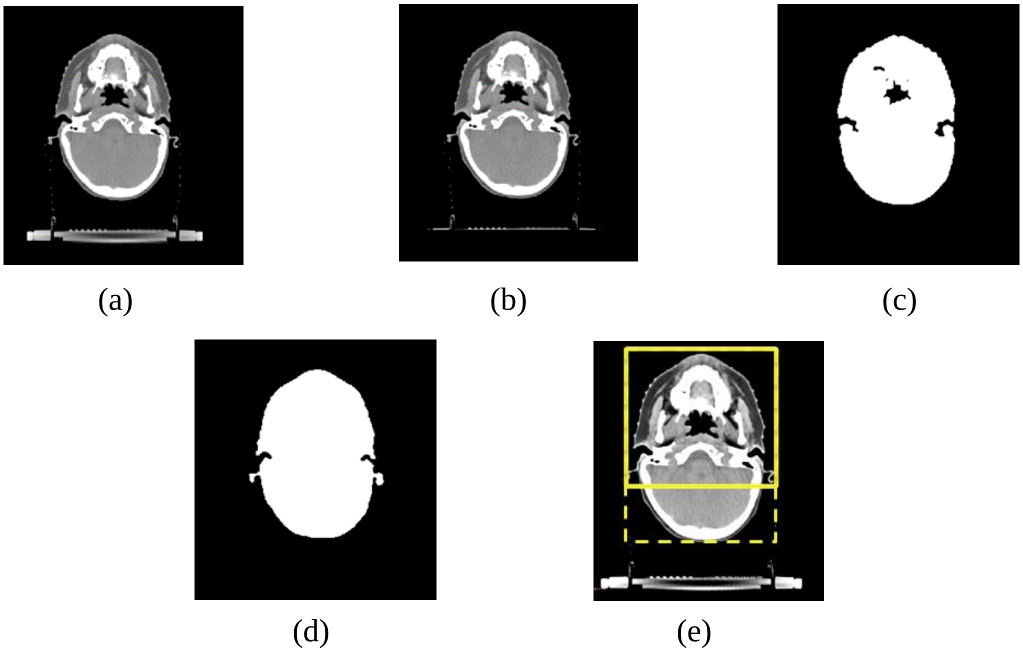


Figure 2. Illustration of method to localize chewing structures in axial H&N CT scans. (a) Input CT image (b) Hough transform-based couch segmentation (c) Intensity thresholding to extract patient outline (d) Morphological post-processing (e) Bounding box with reduced posterior extent.

Model weights were optimized using Stochastic Gradient Descent (SGD) with hyperparameters listed in Table 2 and learning rate was decayed following the polynomial scheduling policy. An early stopping strategy was employed to avoid overfitting if there was no improvement in the validation loss.

2.4 Swallowing structures

2.4.1 Larynx

CT scans were cropped further, with the anterior, left, right, and superior limits defined by the corresponding extents of the previously segmented chewing structures, as shown in figure 3. 2D axial, sagittal and coronal images and corresponding masks of the larynx were extracted within the resulting bounding box and optimization was performed using SGD.

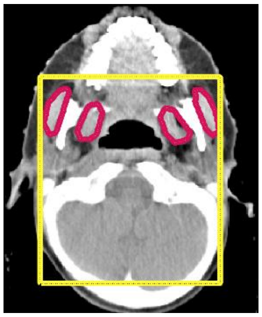


Figure 3. Bounding box (yellow) for localization of larynx using previously segmented chewing structures (red).

2.4.2 Constrictor muscle

To localize the constrictor muscle, CT scans were cropped with the anterior, left, right, and superior limits defined by the corresponding extents of the chewing structures. The posterior and inferior limits were defined by corresponding extents of the larynx, with sufficient padding. 2D axial, sagittal, and coronal images and corresponding masks of the constrictor muscle were extracted within the resulting bounding box, shown in figure 4. Optimization was performed using adaptive moment estimation (Adam) [22].

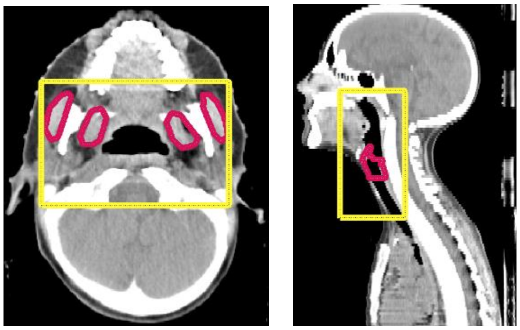


Figure 4. Bounding box (yellow) for localization of constrictor muscle using previously segmented chewing structures and larynx (red).

For each of the above OARs, probability maps returned by the three models (axial, sagittal, and coronal) were averaged and voxels were assigned to the class with the highest resulting probability to produce stable consensus segmentations, robust to occasional failures of the individual models. The segmentation masks were further post-processed to remove isolated voxels by discarding all but the largest connected component. Morphological processing was performed to fill holes and obtain smooth contours.

Table 2. Summary of hyperparameters for training.

Model	Structure(s)	Learning rate	Optimizer	Momentum	Weight Decay
1	MML[a], MMR[b], PML[c], PMR[d]	Axial: 0.002	SGD	0.9	0.0001
		Sagittal: 0.003			
		Coronal: 0.002			

2	Larynx	Axial: 0.0003	SGD	0.9	0.0002
		Sagittal: 0.0003			
		Coronal: 0.0003			
3	CM[e]	Axial: $1\text{x}10^{-6}$	Adam	0.9	0.0001
		Sagittal: $8\text{x}10^{-7}$			
		Coronal : $2\text{x}10^{-6}$			

[a] Masseter (left); [b] masseter (right); [c] medial pterygoid (left); [d] medial pterygoid (right); [e] constrictor muscle.

3. RESULTS

3.1 Comparison to reference standard

The performance of the proposed method was evaluated on the test set of 24 H&N cancer patients by comparing the results against existing manually delineated segmentations. The Dice Similarity Coefficient (DSC) was used to assess the degree of overlap between manual (A) and automated (B) segmentations, computed as:

$$DSC = 2\ \frac{|A \cap B|}{|A| + |B|}$$

Additionally, the 95th percentile of the Hausdorff distance (HD_{95}), i.e., maximum distance between boundary points of A and B, was computed to capture the impact of a few sizeable segmentation errors on the overall segmentation quality. Both DSC and HD_{95} were measured for the individual models (axial, sagittal, coronal) as well as the ensemble to investigate the benefits of a multi-view ensemble.

Examples of segmentations resulting from our algorithm are presented in figure 5 for qualitative assessment. Box plots of DSC and HD_{95} are presented in figure 6.

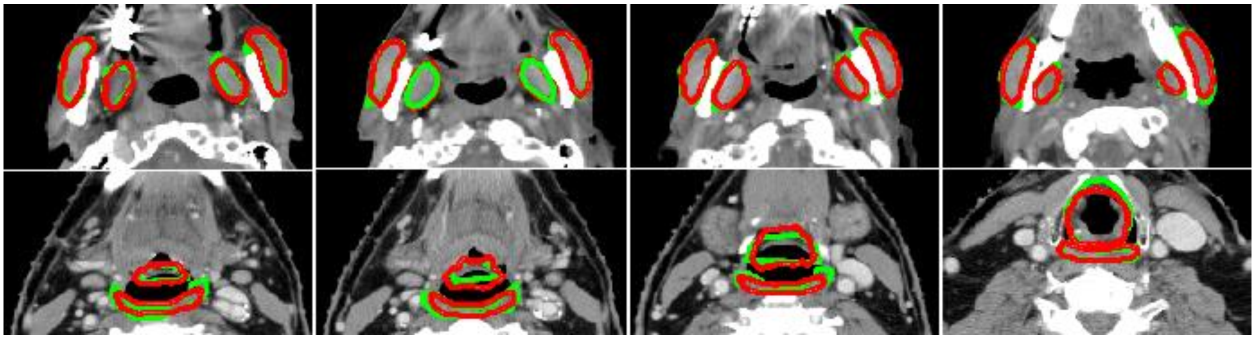


Figure 5. Auto-segmentation results for chewing structures (row-1) and swallowing structures (row-2), shown in four axial cross-sections. Manual reference segmentations are depicted in green and deep-learning-based auto-segmentations in red.

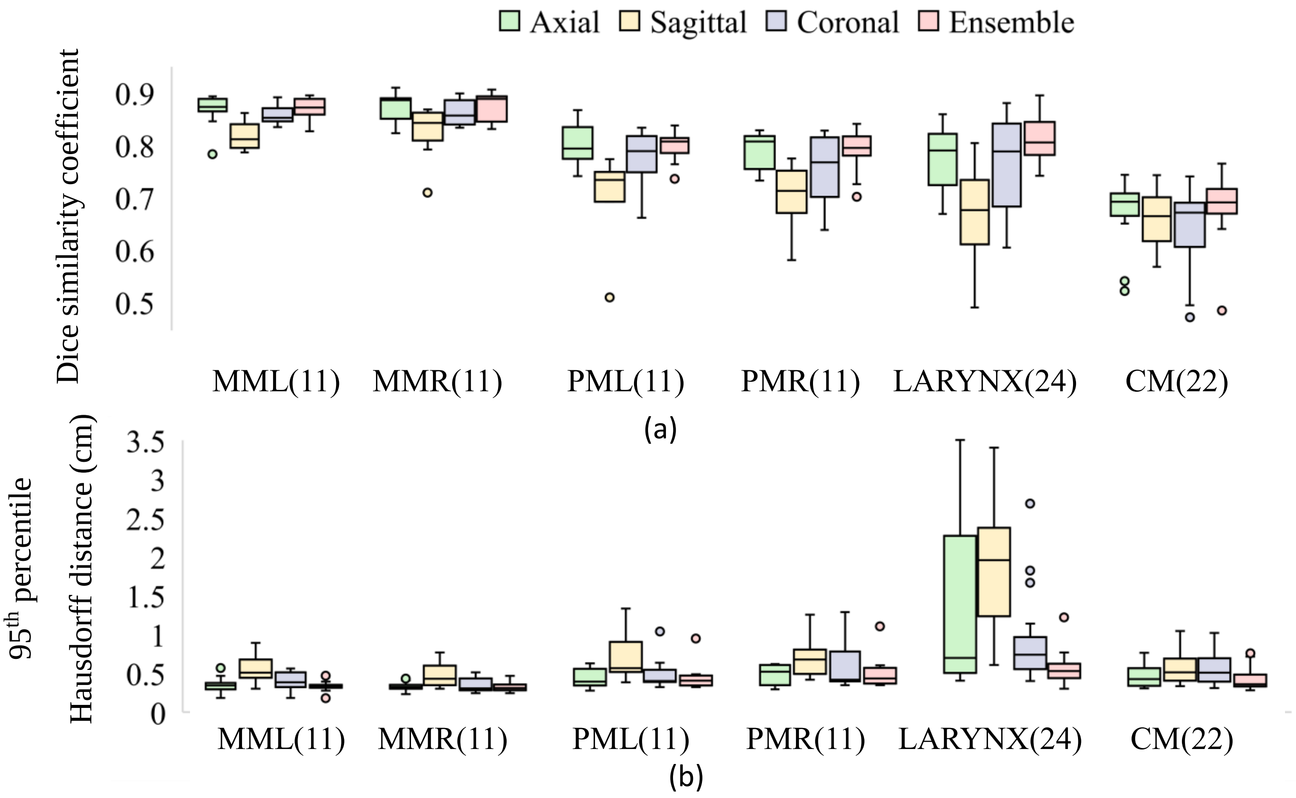


Figure 6. Performance of deep learning models (axial, sagittal, coronal, and ensemble) compared to manual reference segmentations in terms of (a) DSC and (b) HD_{95}. Of the 24 test patients, the number with manual contours available for comparison is noted in parentheses. *MML, MMR: masseters (left and right), PML, PMR: medial pterygoids (left and right), CM : constrictor muscle.*

Differences in mean dose, previously identified [29-30] as a possible factor in radiation-induced complications were also computed between deep learning-based and manual contours, and the Wilcoxon signed-rank test was applied to investigate potential statistical disparities (Table 3). Differences in mean dose were not found to be statistically significant for all tested structures at significance level 5%.

Table 3. Comparison of mean doses extracted using manual and auto-generated contours. Median, first and third quartiles of percentage differences are presented.

Structure	Metric	Difference (%)	p-value	No. test patients
MML[b] U MMR[c]	Ipsilateral[a] mean dose	-0.01 (-1.20, 1.32)	1.00	11
PML[d] U PMR[e]	Ipsilateral[a] mean dose	0.53 (-0.77, 1.11)	0.70	11
Larynx	Mean dose	0.38 (-2.02, 6.59)	0.33	24
CM[f]	Mean dose	0.15 (-0.49, 1.28)	0.29	22

[a] Ipsilaterality for the paired structures was decided based upon the side with the highest dose. [b] Masseter (left); [c] masseter (right); [d] medial pterygoid (left); [e] medial pterygoid (right); [f] constrictor muscle.

3.2 Variability of reference standards

The manual reference segmentations came from various sources: the larynx was delineated for treatment planning by multiple observers with variation in level of expertise; the constrictors and chewing structures were each delineated post-treatment by one of two radiation oncology residents for studying dysphagia and trismus, respectively [29] [30]. Regardless of origin, these contours are referred to as observer-1. A randomly selected subset of 10 CT scans from the testing dataset were then re-segmented by a medical physicist (observer-2) to provide a second reference standard. The agreement between the independent observers was measured in terms of DSC and compared to agreement of each observer with the deep learning-based contours (figure 7). Automated segmentations showed better agreement with each observer as compared to the inter-observer agreement.

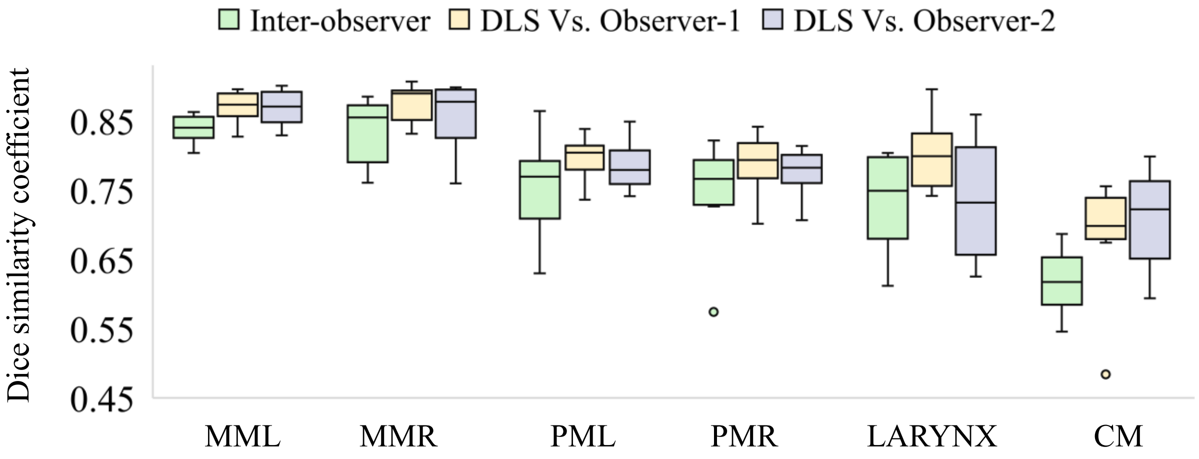


Figure 7. Comparing agreement of deep learning-based segmentations (DLS) with independent observers vs. inter-observer agreement. *MML, MMR: Masseters (left and right), PML, PMR: medial pterygoids (left and right), CM (constrictor muscle)*

3.3 Comparison to previously reported methods

We compared the performance of our method with H&N OARs segmentation models from the published literature. These included state-of-the-art CNN models such as the 3D Unet [13], FocusNet [18], a H&N OAR-focused CNN [12], commercial atlas-based segmentation tool SPICE [5] and other atlas-based methods [24,25]. The mean DSCs for the different methods are summarized in Table 4. The performance of our segmentation models matched or exceeded previously presented methods. However, it should be noted that these results were reported on different datasets and do not represent a direct comparison performed on our testing dataset.

3.4 Distribution of trained models

The segmentation models developed in this work are publicly distributed as a part of CERR's library of model implementations [31]. Model dependencies were encapsulated using Singularity [34] containers, enabling deployment on a variety of scientific computing architectures and aiding in portability and reproducibility. This also facilitates further analysis through integration with CERR's radiomics toolbox [32] and dosimetric models [31]. CERR can be downloaded from https://github.com/cerr/CERR. The list of currently available segmentation containers and corresponding links for download are available at https://github.com/cerr/CERR/wiki/Auto-Segmentation-models. It should be noted that the segmentation models are distributed strictly for research use. Clinical or commercial use is prohibited. CERR and containerized model implementations have not been approved by the U.S. Food and Drug Administration (FDA).

Table 4. DSC (mean ± std. deviation) for swallowing and chewing structures using the proposed method (column-1) and results from previously published methods.

Structure	**Proposed**	**van Rooij et al. [13]** **(2019)** [1]	**Gao et al. [18]** **(2019)** [1]	**Ibragimov et al. [12]** **(2017)** [1]	**Tao et al. [5]** **(2015)** [2]	**Thomson et al. [24]** **(2015)**[2]	**Han et al. [25]** **(2008)**[2]
Masseters	0.87 ± 0.02	-	-	-	-	-	0.83
Pterygoids	0. 80 ± 0.03	-	-	-	-	-	0.83
Larynx	0.81 ± 0.04	0.78 ± 0.05	0.66±0.29	0.86 ± 0.04	0.73 ± 0.04	0.58	-
Constrictor	0.68 ± 0.07	0.68 ± 0.09	-	0.69 ± 0.06	0.65 ± 0.06	0.50	-

[1]Deep learning based [2] atlas based.

4. DISCUSSION

We trained three distinct model ensembles to segment structures of different sizes and constrained the location of each structure group based on the extents of previously identified structures. This sequential localization framework was able to handle imbalance in class labels due to differences in the sizes of the OARs. Additionally, we used an ensemble of models trained on three different image orientations to capture important contextual information and provide redundancy in case of failures of the individual models. The performance of the of the multi-view ensemble in terms of DSC either matched or out-performed the individual models and showed lower variance across all structures, as evidenced by tighter bounds in the box plots (figure 6). Further, the multi-view consensus reduced the worst-case segmentation errors (as captured by HD_{95}) across all the structures. This suggests that employing a multi-view ensemble may help produce more stable segmentations than single-view models

Of the structures considered, the pharyngeal constrictor muscle was most challenging to segment due to its morphological complexity, high anatomical variability, and low soft tissue contrast. This was reflected in our analysis of manual segmentations by different observers. The lowest inter-observer agreement was observed in delineating the constrictor. In some cases, segmentation was further complicated by tumor infiltration around the periphery. For the larynx, standardizing the reference segmentations (generated by multiple observers with varying levels of experience) prior to training could potentially yield better results. We also observed a few spurious and discontinuous detections of the larynx when using models trained only on axial or sagittal images. This was mitigated by generating a consensus segmentation using information from all 3 views (e.g. figure 8). Finally, the deep learning models for all structures were found to generalize well as the auto-generated results showed good agreement with delineations by a new (unseen) observer.

We further investigated the potential for clinical application of the trained models by comparing mean doses extracted from manual and automated segmentations. No statistical differences were observed at the 5% significance level.

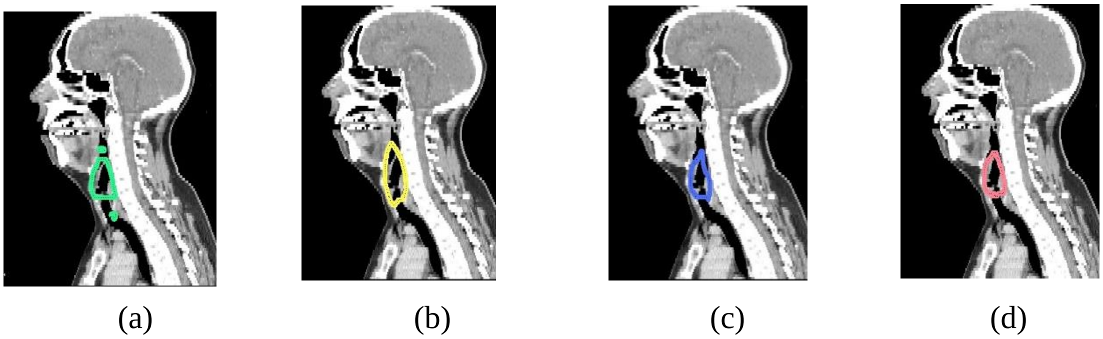


Figure 8. Sagittal cross sections showing auto-segmented larynx using (a) axial only (b) sagittal only (c) coronal only and (d) ensemble models.

5. CONCLUSIONS

We developed a fully-automatic, accurate, and time-efficient method to segment swallowing and chewing structures in CT images and demonstrated its potential for clinical use. The proposed method introduces a novel framework for sequential localization and segmentation to handle imbalances in OAR sizes. Additionally, the potential advantage of a multi-view ensemble over a single-view model was investigated. As hypothesized, the ensemble models were found to yield more stable segmentations across all structures. This ensemble approach could be applied to improve segmentation quality in other sites as well. The trained models are publicly distributed for research use through the open-source platform CERR using Singularity containers.

REFERENCES

1. Harari PM, Song S, Tomé WA. Emphasizing conformal avoidance versus target definition for IMRT planning in head-and-neck cancer. International journal of radiation oncology, biology, physics. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2905233/.
2. Choi M, Refaat T, Lester MS, Bacchus I, Rademaker AW, Mittal BB. Development of a standardized method for contouring the larynx and its substructures. Radiation oncology (London, England). https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4266916/.
3. Levendag PC, Teguh DN, Voet P, et al. Dysphagia disorders in patients with cancer of the oropharynx are significantly affected by the radiation therapy dose to the superior and middle constrictor muscle: A dose-effect relationship. Radiotherapy and Oncology. 2007;85(1):64-73. doi:10.1016/j.radonc.2007.07.009
4. Kraaijenga SA, Hamming-Vrieze O, Verheijen S, et al. Radiation dose to the masseter and medial pterygoid muscle in relation to trismus after chemoradiotherapy for advanced head and neck cancer. Head & Neck. 2019;41(5):1387-1394. doi:10.1002/hed.25573
5. Tao C-J, Yi J-L, Chen N-Y, et al. Multi-subject atlas-based auto-segmentation reduces interobserver variation and improves dosimetric parameter consistency for organs at risk in

nasopharyngeal carcinoma: A multi-institution clinical study. Radiotherapy and Oncology. 2015;115(3):407-411. doi:10.1016/j.radonc.2015.05.012

6. Levendag P, Hoogeman M, Teguh D, et al. Atlas Based Auto-segmentation of CT Images: Clinical Evaluation of using Auto-contouring in High-dose, High-precision Radiotherapy of Cancer in the Head and Neck. International Journal of Radiation Oncology*Biology*Physics. 2008;72(1). doi:10.1016/j.ijrobp.2008.06.1285
7. Fortunati V, Verhaart RF, Lijn FVD, et al. Tissue segmentation of head and neck CT images for treatment planning: A multiatlas approach combined with intensity modeling. Medical Physics. 2013;40(7):071905. doi:10.1118/1.4810971
8. Qazi AA, Pekar V, Kim J, Xie J, Breen SL, Jaffray DA. Auto-segmentation of normal and target structures in head and neck CT images: A feature-driven model-based approach. Medical Physics. 2011;38(11):6160-6170. doi:10.1118/1.3654160
9. Wang Z, Wei L, Wang L, Gao Y, Chen W, Shen D. Hierarchical Vertex Regression-Based Segmentation of Head and Neck CT Images for Radiotherapy Planning. IEEE Transactions on Image Processing. 2018;27(2):923-937. doi:10.1109/tip.2017.2768621
10. Fritscher KD, Peroni M, Zaffino P, Spadea MF, Schubert R, Sharp G. Automatic segmentation of head and neck CT images for radiotherapy treatment planning using multiple atlases, statistical appearance models, and geodesic active contours. Medical Physics. 2014;41(5):051910. doi:10.1118/1.4871623
11. Zhu W, Huang Y, Zeng L, et al. AnatomyNet: Deep learning for fast and fully automated whole-volume segmentation of head and neck anatomy. Medical Physics. 2018;46(2):576-589. doi:10.1002/mp.13300

12. Ibragimov B, Xing L. Segmentation of organs-at-risks in head and neck CT images using convolutional neural networks. Medical Physics. 2017;44(2):547-557. doi:10.1002/mp.12045

13. Rooij WV, Dahele M, Brandao HR, Delaney AR, Slotman BJ, Verbakel WF. Deep Learning-Based Delineation of Head and Neck Organs at Risk: Geometric and Dosimetric Evaluation. International Journal of Radiation Oncology*Biology*Physics. 2019;104(3):677-684. doi:10.1016/j.ijrobp.2019.02.040

14. Ronneberger O, Fischer P, Brox T. U-Net: Convolutional Networks for Biomedical Image Segmentation. Lecture Notes in Computer Science Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015. 2015:234-241. doi:10.1007/978-3-319-24574-4_28

15. Men K, Chen X, Zhang Y, et al. Deep Deconvolutional Neural Network for Target Segmentation of Nasopharyngeal Cancer in Planning Computed Tomography Images. Frontiers in Oncology. 2017;7. doi:10.3389/fonc.2017.00315

16. Simonyan K, Zisserman A. Very Deep Convolutional Networks for Large-Scale Image Recognition. https://arxiv.org/abs/1409.1556.

17. Tong N, Gou S, Yang S, Ruan D, Sheng K. Fully automatic multi-organ segmentation for head and neck cancer radiotherapy using shape representation model constrained fully convolutional neural networks. Medical Physics. 2018;45(10):4558-4567. doi:10.1002/mp.13147

18. Gao Y, Huang R, Chen M, et al. FocusNet: Imbalanced Large and Small Organ Segmentation with an End-to-End Deep Neural Network for Head and Neck CT Images. Lecture Notes in Computer Science Medical Image Computing and Computer Assisted Intervention – MICCAI 2019. 2019:829-838. doi:10.1007/978-3-030-32248-9_92

19. Chen L-C, Zhu Y, Papandreou G, Schroff F, Adam H. Encoder-Decoder with Atrous Separable Convolution for Semantic Image Segmentation. Computer Vision – ECCV 2018 Lecture Notes in Computer Science. 2018:833-851. doi:10.1007/978-3-030-01234-2_49

20. He K, Zhang X, Ren S, Sun J. Deep Residual Learning for Image Recognition. 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). 2016. doi:10.1109/cvpr.2016.90

21. Zhang J. jfzhang95/pytorch-deeplab-xception. GitHub. https://github.com/jfzhang95/pytorch-deeplab-xception.git.

22. Kingma D, Ba J. Adam: A Method for Stochastic Optimization. arXiv preprint. https://arxiv.org/pdf/1412.6980.

23. Deasy JO, Blanco AI, Clark VH. CERR: A computational environment for radiotherapy research. Medical Physics. 2003;30(5):979-985. doi:10.1118/1.1568978

24. Thomson D, Boylan C, Liptrot T, et al. Evaluation of an automatic segmentation algorithm for definition of head and neck organs at risk. Radiation Oncology. 2014;9(1):173. doi:10.1186/1748-717x-9-173.

25. Han X, Hoogeman MS, Levendag PC, et al. Atlas-Based Auto-segmentation of Head and Neck CT Images. Medical Image Computing and Computer-Assisted Intervention – MICCAI 2008. 2008:434-441. doi:10.1007/978-3-540-85990-1_52

26. Elguindi S, Zelefsky MJ, Jiang J, et al. Deep learning-based auto-segmentation of targets and organs-at-risk for magnetic resonance imaging only planning of prostate radiotherapy. Physics and Imaging in Radiation Oncology. 2019;12:80-86. doi:10.1016/j.phro.2019.11.006

27. Haq R, Hotca A, Apte A, Rimner A, Deasy JO, Thor M. Cardio-pulmonary substructure segmentation of radiotherapy computed tomography images using convolutional neural

networks for clinical outcomes analysis. Physics and Imaging in Radiation Oncology. https://www.sciencedirect.com/science/article/abs/pii/S2405631620300221

28. Folk, M., Cheng, A., & Yates, K. (1999, November). HDF5: A file format and I/O library for high performance computing applications. In Proceedings of supercomputing (Vol. 99, pp. 5-33).

29. Tsai CJ, Jackson A, Setton J, et al. Modeling Dose Response for Late Dysphagia in Patients With Head and Neck Cancer in the Modern Era of Definitive Chemoradiation. JCO Clinical Cancer Informatics. 2017;(1):1-7. doi:10.1200/cci.17.00070

30. Rao SD, Saleh ZH, Setton J, et al. Dose-volume factors correlating with trismus following chemoradiation for head and neck cancer. Acta Oncologica. 2015;55(1):99-104. doi:10.3109/0284186x.2015.1037864

31. Apte AP, Iyer A, Thor M, et al. Library of deep-learning image segmentation and outcomes model-implementations. Physica Medica. https://www.sciencedirect.com/science/article/pii/S1120179720300922. Published May 1, 2020. Accessed July 27, 2020.

32. Apte AP, Iyer A, Crispin-Ortuzar M, et al. Technical Note: Extension of CERR for computational radiomics: A comprehensive MATLAB platform for reproducible radiomics research. Medical Physics. 2018;45(8):3713-3720. doi:10.1002/mp.13046

33. Paszke, A.; Gross, S.; Massa, F.; Lerer, A.; Bradbury, J.; Chanan, G.; Killeen, T.; Lin, Z.; Gimelshein, N.; Antiga, L.; et al. PyTorch: An imperative style, high-performance deep learning library. In Proceedings of the Advances in Neural Information Processing Systems, Vancouver, BC, Canada, 8–14 December 2019; pp. 8024–8035.

34. Kurtzer GM, Sochat V, Bauer MW. Singularity: Scientific containers for mobility of compute. Plos One. 2017;12(5). doi:10.1371/journal.pone.0177459

35. Iyer A, Thor M, Haq R, Deasy JO, Apte AP. Deep learning-based auto-segmentation of swallowing and chewing structures in CT. bioRxiv. https://www.biorxiv.org/content/10.1101/772178v1. Published January 1, 2019. Accessed July 27, 2020.

ACKNOWLEDGEMENTS

This work was partially funded by NIH grant 1R01CA198121 and NIH/NCI Cancer Center Support grant P30 CA008748.