Detegmentation of Organs-at-Risk for Head and Neck Cancer

Shan Wang¹ Jax Luo² Xu Qiao¹ Nazim Haouchine³ Scott Raymond²

202300172023@MAIL.SDU.EDU.CN JLUO5@MGH.HARVARD.EDU QIAOXU@SDU.EDU.CN NHAOUCHINE@BWH.HARVARD.EDU RAYMONS3@CCF.ORG

¹ School of Control Science and Engineering, Shandong University, Shandong, China

² Neurological Institute, Cleveland Clinic, OH, USA

³ Brigham and Women's Hospital, Harvard Medical School, MA, USA

Editors: Accepted for publication at MIDL 2025

Abstract

Segmentation of organs-at-risk (OARs) is a critical step in radiation therapy planning for head and neck cancer (HNC). Recent advancements in fine-tuned foundation models have demonstrated the potential to surpass conventional segmentation methods when provided with appropriately sized box prompts. We present a practical Detegmentation framework that integrates a detection network to autonomously generate box prompts for training and testing a foundation model for OARs segmentation in HNC. Our approach outperforms leading methods from the most recent HaN-Seg Challenge, highlighting its strong potential for clinical application.

Keywords: Detegmentation, Detection, Segmentation, Foundational Model, Head and Neck Cancer

1. Introduction

Head and neck (HNC) cancer ranks among the most prevalent malignancies worldwide. In radiation therapy for HNC, accurate segmentation of organs at risk (OARs) is critical to minimizing post-treatment radiation-induced complications (Ye et al., 2022). Recent advancements in fine-tuned foundation models have demonstrated the potential to surpass conventional segmentation methods when provided with appropriately sized box prompts—rectangular regions that indicate the location of the target object (Kirillov et al., 2023; Xu et al., 2024, 2025; Fan et al., 2019; Chen et al., 2024). However, manually delineating these boxes for OARs is impractical, as it requires substantial clinician input and is susceptible to intra- and inter-rater variability. In this study, we propose a Detegmentation framework that integrates a detection network to autonomously generate fitted box prompts for each target OAR, enabling streamlined training and testing of a foundation segmentation model without clinician intervention. Our approach outperforms leading methods from the most recent HaN-Seg Challenge(Podobnik et al., 2023), highlighting its strong potential for clinical application.



Figure 1: An overview of the proposed Detegmentation framework. D represents the incorporated detection model.

2. Methods

The Detegmentation framework comprises an nnDetect(Baumgartner et al., 2021) model for object detection and a 3D adaptation of Segment Anything Model (SAM)(Chen et al., 2024), which can be fine-tuned(Hu et al., 2021) for specific OARs segmentation. As illustrated in Fig. 1, we first train the nnDetect model D to automatically predict the locations and bounding box coordinates of the target OARs. These coordinates are then used as box prompts to fine-tune the 3D SAM for enhanced OAR segmentation. During inference, the test case is first processed by D to generate the corresponding box prompt. This prompt, together with the input image, is then fed into the segmentation model to produce the predicted segmentation label.

3. Experiments

We assessed the proposed framework using the brainstem and left parotid gland structures from the publicly available HaN-Seg challenge dataset(Podobnik et al., 2023), which was partitioned into 30 cases for training and 12 cases for testing.

Our method demonstrates reliable detection performance, achieving a sensitivity of 1.0 for both the brainstem and the left parotid gland. The Dice similarity coefficient between the ground truth and predicted bounding boxes is 0.78. To ensure complete coverage of the target structures during segmentation, we expanded the input bounding box by 5 voxels in each direction.

As summarized in Table.1, the proposed Detegmentation framework yields a mean Dice score of 0.895 and an HD95 of 1.8 mm for brainstem segmentation, and a Dice score of 0.886 with an HD95 of 2.0 mm for left parotid gland segmentation, surpassing the top performing nnU-Net(Isensee et al., 2018) and DynUNet(Yang et al., 2025) methods in the latest HaN-seg challenge. Fig. 2 presents qualitative results for the segmented brainstem and left parotid gland.

4. Discussion and Conclusion

We present a practical Detegmentation framework that integrates a detection network to autonomously generate box prompts for training and testing a foundation model for Organs at Risk segmentation in Head and Neck cancer. Our method enables accurate OAR

 Table 1: Quantitave segmentation results comparing the proposed Detegmentation framework with leading methods in the HaN-seg challenge

| OAR | Mean | nnU-Net | DynUNet | Ours |
|------------------|-------------|--|---|---|
| Brainstem | DSC HD95 | $\begin{array}{c} 0.885\\ 3.9 \end{array}$ | $\begin{array}{c} 0.849 \\ 4.7 \end{array}$ | $\begin{array}{c} 0.895\\ 1.8\end{array}$ |
| Parotid gland(L) | DSC HD95 | $0.867 \\ 5.1$ | $0.851 \\ 5.1$ | $\begin{array}{r} 0.886\\ 2.0\end{array}$ |



Figure 2: Qualitative results showcasing the predicted segmented brainstem and the left parotid gland labels, along with their 3D renderings. GT denotes the ground truth label.

segmentation without clinician intervention, highlighting its strong potential for clinical application.

Future research will aim to extend the proposed framework to encompass a broader range of OARs, with particular emphasis on smaller and more challenging structures such as the optic chiasm and optic nerves. Additionally, more advanced object detection models will be explored to address instances where the current nnDetect model underperforms. Another important direction involves integrating multiple imaging modalities (e.g., CT and PET) as input, with the objective of further enhancing segmentation accuracy.

References

Michael Baumgartner, Paul F Jäger, Fabian Isensee, and Klaus H Maier-Hein. nndetection: a self-configuring method for medical object detection. In *Medical Image Computing and Computer Assisted Intervention–MICCAI 2021: 24th International Conference, Stras-* bourg, France, September 27–October 1, 2021, Proceedings, Part V 24, pages 530–539. Springer, 2021.

- Cheng Chen, Juzheng Miao, Dufan Wu, Aoxiao Zhong, Zhiling Yan, Sekeun Kim, Jiang Hu, Zhengliang Liu, Lichao Sun, Xiang Li, et al. Ma-sam: Modality-agnostic sam adaptation for 3d medical image segmentation. *Medical Image Analysis*, 98:103310, 2024.
- Guoxin Fan, Huaqing Liu, Zhenhua Wu, Yufeng Li, Chaobo Feng, Dongdong Wang, Jie Luo, WM Wells, and Shisheng He. Deep learning–based automatic segmentation of lumbosacral nerves on ct for spinal intervention: a translational study. *American Journal of Neuroradiology*, 40(6):1074–1081, 2019.
- Edward J Hu, Yelong Shen, Phillip Wallis, Zeyuan Allen-Zhu, Yuanzhi Li, Shean Wang, Lu Wang, and Weizhu Chen. Lora: Low-rank adaptation of large language models. *arXiv* preprint arXiv:2106.09685, 2021.
- Fabian Isensee, Jens Petersen, Andre Klein, David Zimmerer, Paul F Jaeger, Simon Kohl, Jakob Wasserthal, Gregor Koehler, Tobias Norajitra, Sebastian Wirkert, et al. nnu-net: Self-adapting framework for u-net-based medical image segmentation. arXiv preprint arXiv:1809.10486, 2018.
- Alexander Kirillov, Eric Mintun, Nikhila Ravi, Hanzi Mao, Chloe Rolland, Laura Gustafson, Tete Xiao, Spencer Whitehead, Alexander C Berg, Wan-Yen Lo, et al. Segment anything. In Proceedings of the IEEE/CVF international conference on computer vision, pages 4015–4026, 2023.
- Gašper Podobnik, Primož Strojan, Primož Peterlin, Bulat Ibragimov, and Tomaž Vrtovec. Han-seg: The head and neck organ-at-risk ct and mr segmentation dataset. *Medical physics*, 50(3):1917–1927, 2023.
- Guoping Xu, Xiaoxue Qian, Hua-Chieh Shao, Jax Luo, Weiguo Lu, and You Zhang. A segment anything model-guided and match-based semi-supervised segmentation framework for medical imaging. *Medical Physics*, 2025.
- Zhe Xu, Donghuan Lu, Jie Luo, Yefeng Zheng, and Raymond Kai-yu Tong. Separated collaborative learning for semi-supervised prostate segmentation with multi-site heterogeneous unlabeled mri data. *Medical Image Analysis*, 93:103095, 2024.
- Jin Yang, Daniel S Marcus, and Aristeidis Sotiras. Dynamic u-net: adaptively calibrate features for abdominal multiorgan segmentation. In *Medical Imaging 2025: Computer-Aided Diagnosis*, volume 13407, pages 326–334. SPIE, 2025.
- Xianghua Ye, Dazhou Guo, Jia Ge, Senxiang Yan, Yi Xin, Yuchen Song, Yongheng Yan, Bing-shen Huang, Tsung-Min Hung, Zhuotun Zhu, et al. Comprehensive and clinically accurate head and neck cancer organs-at-risk delineation on a multi-institutional study. *Nature communications*, 13(1):6137, 2022.