



Comparison of Deep Learning-Based Auto-Segmentation Algorithms for Head and Neck Cancer

Min Seo CHOI, Jaehee CHUN, Nalee Kim, Ji Suk CHANG, Jinsung KIM
Department of Radiation Oncology, Yonsei Cancer Center, Yonsei University College of Medicine, 50-1 Yonsei-ro, Seodaemun-gu, Seoul, 03722, South Korea

Introduction

Accurate delineation of the organs at risk and targets is a crucial part of head and neck (H&N) cancer radiation therapy planning. Increased toxicity to the surrounding healthy tissues must be avoided in order to reduce the likelihood of side effects such as xerostomia and mucositis which can severely affect the patients' quality of life following the treatment [1].

However, it is well-known that manually segmenting normal and target structures in this site is especially challenging due to a large number of OARs near H&N tumors, making it very laborious and subject to inter-observer variations [2].

In order to overcome this issue and the need for manual segmentation, many previous studies on H&N auto-segmentation have been conducted [3]–[7]. The aim of our study is to compare and evaluate the state-of-the-art deep learning based auto-segmentation (DBAS) approaches.

Methods

The CT datasets used in this study consisted of 115 simulation CT scans of head and neck cancer patients and manual delineations of ten organs at risk (OAR) drawn by a single expert: brainstem, spinal cord, parotids, oral cavity, submandibular glands, thyroid, mandible, and esophagus. The data was split into 80 training sets, 10 validation sets, and 25 test sets.

Modified full solution licenseNet (FC-DenseNet) was used for DBAS development (Figure a). The existing 2D has been expanded from 2D to 3D structure, and hyper-parameters such as ground rate, dropout rate, batch size and layers per block have been changed to use maximum GPU memory in various experimental environments.

In this study, we also included two additional commercial DBAS software packages: AccuContour (Manteia tech, Xiamen, China) and DLCExpert (Mirada Medical, Oxford, United Kingdom). These packages were pre-trained by the manufacturer and hence did not utilize our training datasets, and were only used for evaluation and comparison purposes. Lastly, the Dice similarity coefficient (DSC) was used to compare the similarity of the auto-segmentation to the manually expert generated contours.

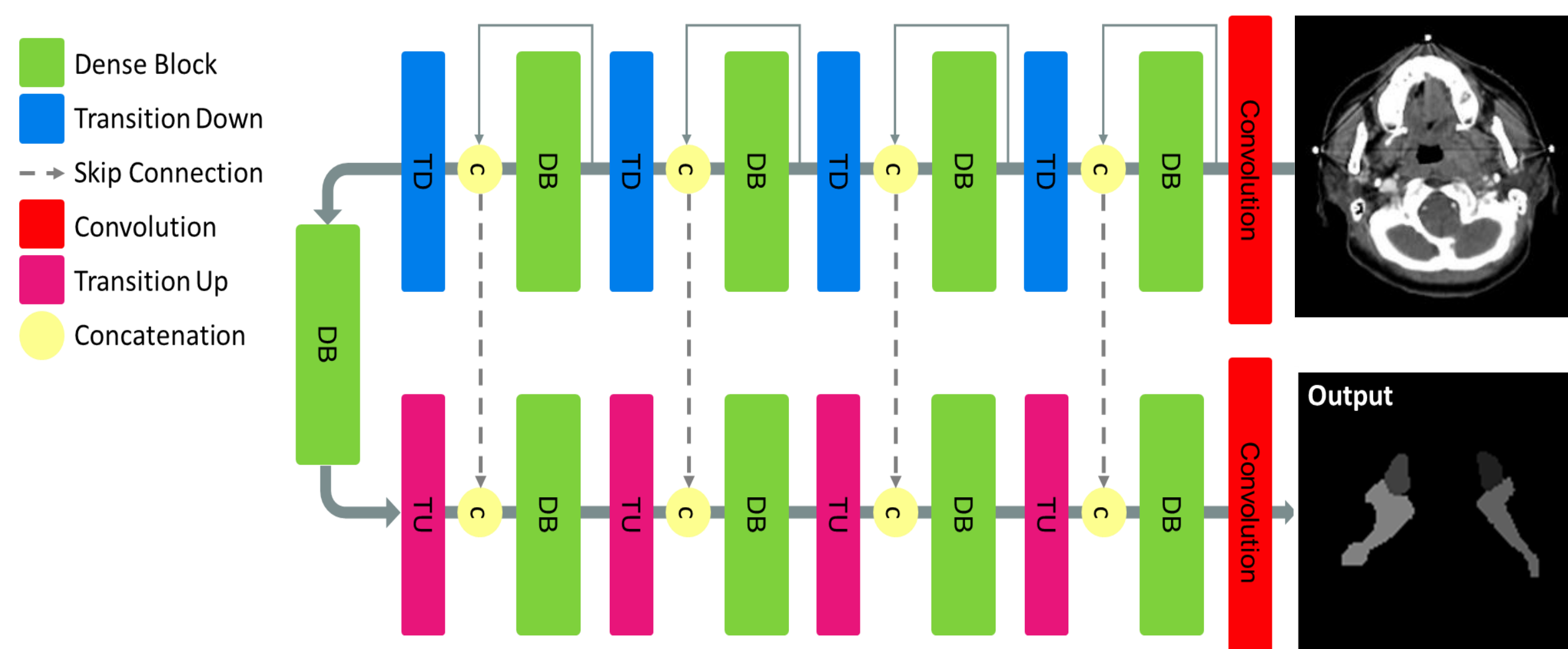


Figure (a) : The schematic of the proposed FC-DenseNet (FCDN) architecture

Results

The segmentation results of each software and their DSC scores are shown in Figures b and c. Our in-house software's DSCs ranged from 81% similarity in the esophagus and up to 95% similarity in the mandible and produced the highest average DSCs in every OARs except for the right submandibular gland and esophagus.

The two commercial DBAS packages showed an average DSC of over 70% across most structures. AccuContour outperformed DLCExpert in eight out of 10 structures. DLCExpert showed limited performance in the spinal cord and esophagus respectively, but this is likely to be the result of being trained on an external dataset with different structure definitions, possibly with different starting and ending points.

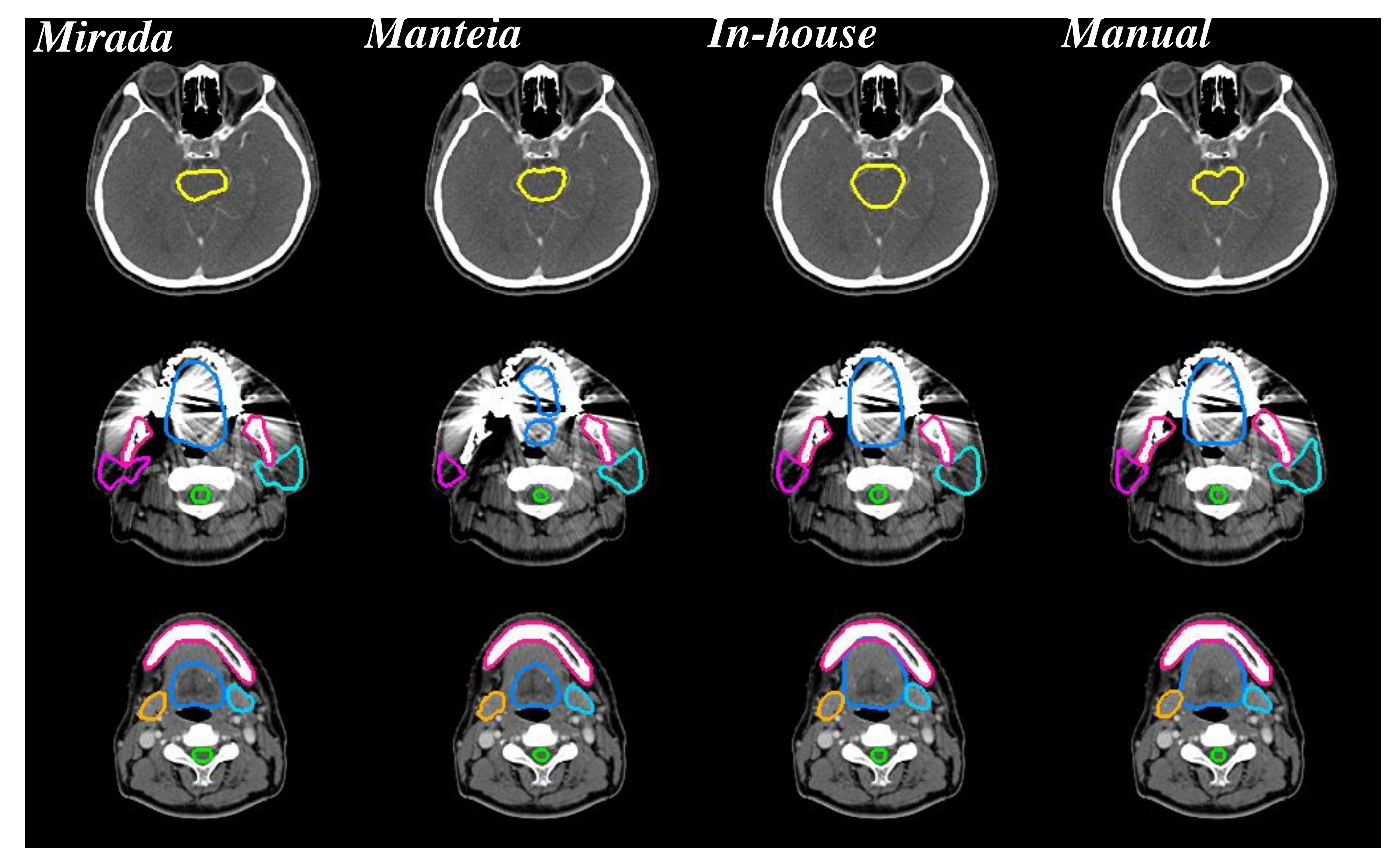


Figure (b) : Examples of the auto-segmentation generated from each software packages compared against the ground-truth manual contours

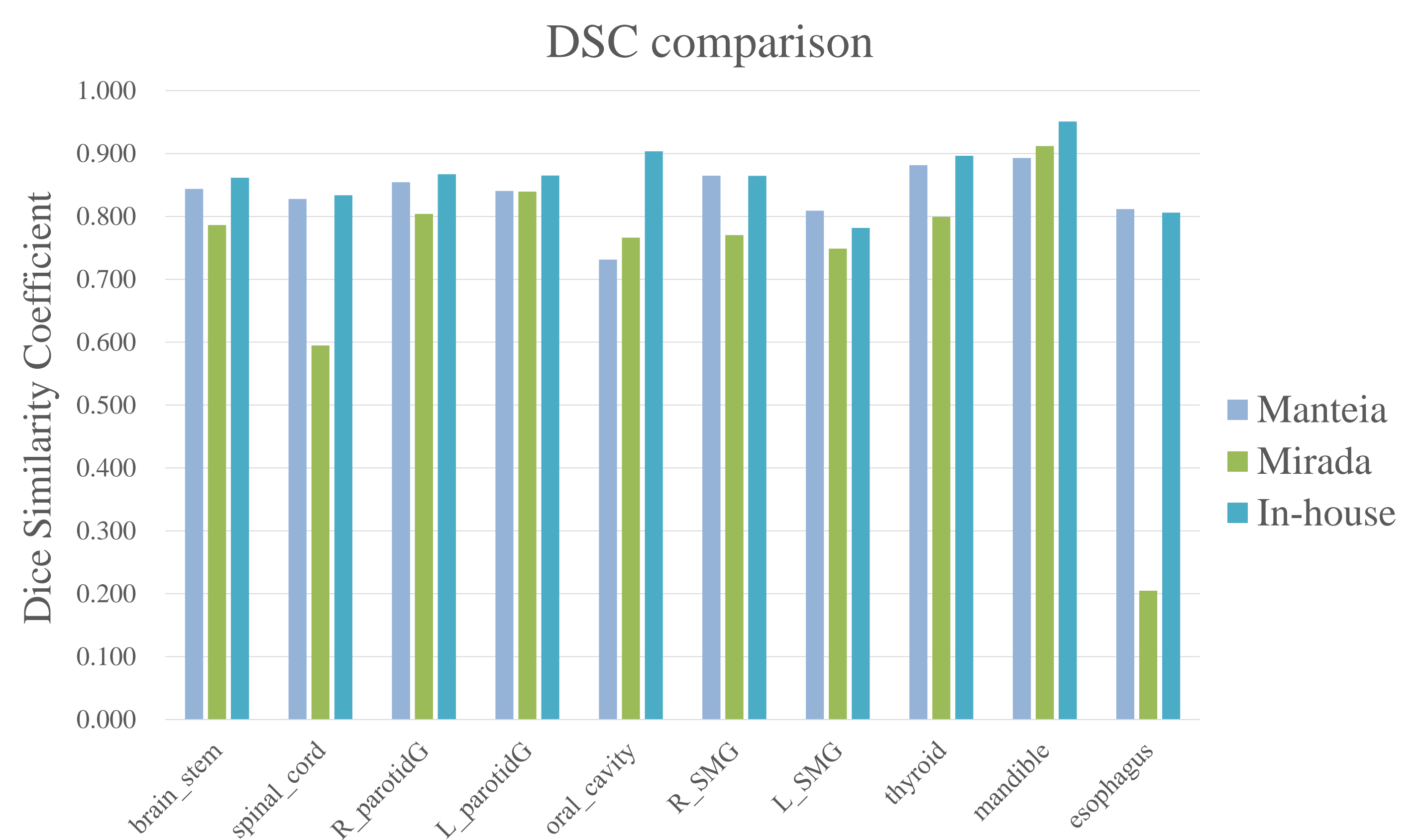


Figure (c) : The Dice similarity comparison between DBAS software

Conclusion / Future Works

In summary, we have compared multiple DBAS algorithms for H&N cancer segmentation for radiotherapy planning for the first time. Through similarity evaluation, we have found that our in-house model produced the highest similarity to the ground truth which implies our model mimicked the drawing style of the expert the best out of the three algorithms.

Our results also highlight the importance of on-site training with institution-specific datasets to ensure better segmentation outcomes from a DBAS algorithms.

REFERENCES

- [1] P. Giraud et al., "Radiomics and machine learning for radiotherapy in head and neck cancers," *Frontiers in Oncology*, 2019, doi: 10.3389/fonc.2019.00174.
- [2] C. E. Cardenas et al., "Deep Learning Algorithm for Auto-Delineation of High-Risk Oropharyngeal Clinical Target Volumes With Built-In Dice Similarity Coefficient Parameter Optimization Function," *Int. J. Radiat. Oncol. Biol. Phys.*, vol. 101, no. 2, pp. 468–478, 2018, doi: 10.1016/j.ijrobp.2018.01.114.
- [3] J. Wong et al., "Comparing deep learning-based auto-segmentation of organs at risk and clinical target volumes to expert inter-observer variability in radiotherapy planning," *Radiother. Oncol.*, vol. 144, pp. 152–158, 2020, doi: 10.1016/j.radonc.2019.10.019.
- [4] K. Men et al., "Technical Note: More accurate and efficient segmentation of organs-at-risk in radiotherapy with convolutional neural networks cascades," *Med. Phys.*, vol. 46, no. 1, pp. 286–292, 2019, doi: 10.1002/mp.13296.
- [5] S. Nikolov et al., "Deep learning to achieve clinically applicable segmentation of head and neck anatomy for radiotherapy," pp. 1–31, 2018, [Online]. Available: <http://arxiv.org/abs/1809.04430>.
- [6] L. V. van Dijk et al., "Improving automatic delineation for head and neck organs at risk by Deep Learning Contouring," *Radiother. Oncol.*, vol. 142, pp. 115–123, 2020, doi: 10.1016/j.radonc.2019.09.022.
- [7] J. van der Veen et al., "Benefits of deep learning for delineation of organs at risk in head and neck cancer," *Radiother. Oncol.*, vol. 138, pp. 68–74, Sep. 2019, doi: 10.1016/j.radonc.2019.05.010.