Comparison of Multi-Organ Segmentation Tools for Whole-Body [^{18}F]FDG-PET/CT Clinical Imaging

Cameron Wheeler[1,2,3], Phyo H. Khaing[2], Eleonora D'Arnese[1], Adriana A.S. Tavares[2,3]

(1) School of Informatics, University of Edinburgh, Edinburgh, Scotland, United Kingdom.

(2) Centre for Cardiovascular Science, University of Edinburgh, Edinburgh, Scotland, United Kingdom.

(3) Edinburgh Imaging, University of Edinburgh, Edinburgh, Scotland, United Kingdom.

Typically, delineation of volumes of interest (VOI) within clinical PET/CT imaging is performed manually. This is time consuming [1] and is highly vulnerable to inter/intra-operator variability resulting in differing delineations for the same patient [1]. Application of automated segmentation aims to address these issues in the VOI delineation process; increasing image throughput and reducing the current manual stochasticity involved. Moreover, automation in the segmentation process has the potential to galvanise downstream analyses, including network analysis [2]. Among the various automated multi-organ segmentation approaches, two methodologies Multiple-Organ Objective Segmentation (MOOSE) [3] and TotalSegmentator (TS) [4] have emerged as state-of-the-art (SOTA). Both methods leverage the nnU-Net framework as their underlying architecture. However, they differ in training dataset, nnU-Net configurations, and weights. Recently, Julie et al. [5] compared MOOSE and TotalSegmentator on a metastatic breast cancer dataset. However, gold-standard labels were not generated for this study and as a result, comparison against a verified gold-standard is not available, instead the authors focus on the degree of agreement and differences in feature values.

Concurrent with Julie et al. [5], we look to compare MOOSE and TS, on a dataset of clinical stage IIB/III non-small cell lung carcinoma (NSCLC). We compare both methods versus gold-standard manual delineation and evaluate using current technical segmentation metrics such as Dice-Sørensen coefficient (DSC) alongside PET/CT outcome metrics (e.g. Hounsfield units (HU) and Standardised Uptake Values (SUV)) to assess if the automated methods introduce quantitative bias versus manual annotations.

Technical evaluation found comparable performance between TS and MOOSE, with similar DSC for thoracic organs (Left Lung: 0.84 TS vs. 0.85 MOOSE; Right Lung: 0.83 vs. 0.83; Left Kidney: 0.70 vs. 0.70; Right Kidney: 0.70 vs. 0.69; Liver: 0.85 vs. 0.85). However, clinical endpoint analyses revealed more nuanced differences. Significant differences were observed in SUVmax values between TS and MOOSE, Left Lung ($p = 0.012$) and Right Lung ($p = 0.011$). These findings show that while technical evaluation suggests the methods are largely interchangeable, clinical analyses highlight subtle yet important behavioural

differences, particularly in the segmentation of pathological organs. Furthermore, we found significant errors in cerebral segmentation, TS achieved a DSC of 0.87 compared to 0.67 for MOOSE. Together, our results describe the importance of analysing automated segmentation tools using both technical and clinical points of view and outlines future improvements for the methodologies.

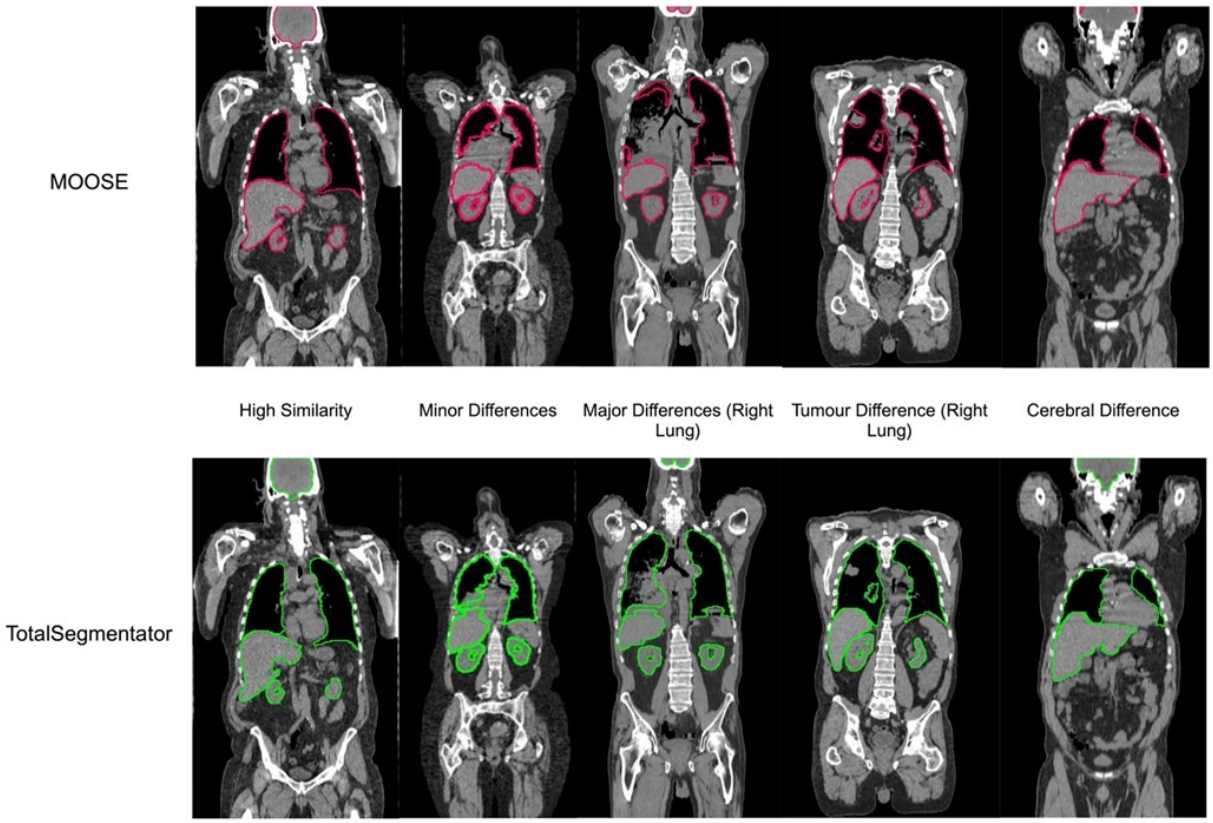

Figure. 1 Comparison of segmentation results from MOOSE and TotalSegmentator: While many cases exhibited high agreement between the two methods, others showed minor to substantial differences. Notably, discrepancies were observed in inclusion of the tumour within the lung segmentation and segmentation of cerebral regions, which emerged as consistent points of contrast between the two methods.

References:

(1) Qu, C., Zhang, T., Qiao, H., Liu, J., Tang, Y., Yuille, A.L., Zhou, Z.: Abdomenatlas-8k: annotating 8,000 ct volumes for multi-organ segmentation in three weeks. In: Proceedings of the 37th International Conference on Neural Information Processing Systems. NIPS '23. Curran Associates Inc., Red Hook,NY, USA (2024)

(2) Suchacki, K.J., Alcaide-Corral, C.J., Nimale, S., Macaskill, M.G., Stimson, R.H., Farquharson, C., Freeman, T.C., Tavares, A. A.: A systems-level analysis of total-body pet data reveals complex skeletal metabolism networks in vivo. Frontiers in Medicine 8, 740615 (2021)

(3) Sundar, L.K.S., Yu, J., Muzik, O., Kulterer, O.C., Fueger, B., Kifjak, D., Nakuz, T., Shin, H.M., Sima, A.K., Kitzmantl, D., et al.: Fully automated, semantic

segmentation of whole-body 18f-fdg pet/ct images based on data-centric artificial intelligence. Journal of Nuclear Medicine 63(12), 1941–1948 (2022)

(4) Wasserthal, J., Breit, H.-C., Meyer, M.T., Pradella, M., Hinck, D., Sauter, A.W., Heye, T., Boll, D.T., Cyriac, J., Yang, S., et al.: Totalsegmentator: robust segmentation of 104 anatomic structures in ct images. Radiology: Artificial Intelligence 5(5) (2023)

(5) Julie, A., Shiyam Sundar, L.K., Seban, R.-D., Luporsi, M., Nioche, C., Beyer, T., Buvat, I., ORLHAC, F.: Moose vs totalsegmentator: Comparison of feature values of segmented anatomical regions in [18f]fdg pet/ct images. Journal of Nuclear Medicine 65(supplement 2), 241948–241948 (2024)