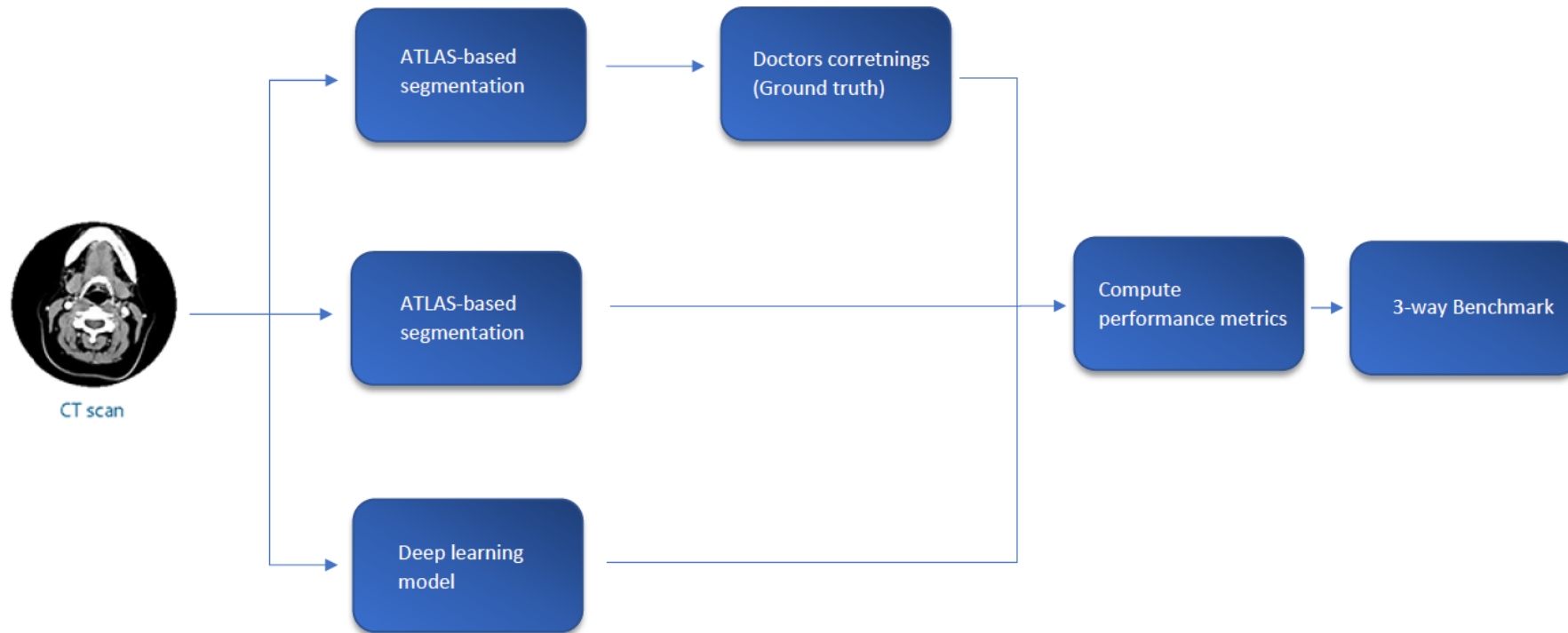


Performance Testing of Deep Learning and Auto- Segmentation Algorithms

Alex Kolby
Eskild H. Andersen
Simon Rydder

Problem description

- ▶ Investigate the difference in performance of autosegmentation algorithms for delineation of organs.
- ▶ Create a good performance measurement for benchmarking.



The data

- ▶ Database consisting of approx 800 patients treated for cancer at Aarhus Universitetshospital in the period 2005-2016
- ▶ Data consists of CT scans and the delineation of organs for each patient.
- ▶ Semi 3D pictures. Consist of slices varying from 2-3 mm. Put together in a stack they imitate a 3D picture.

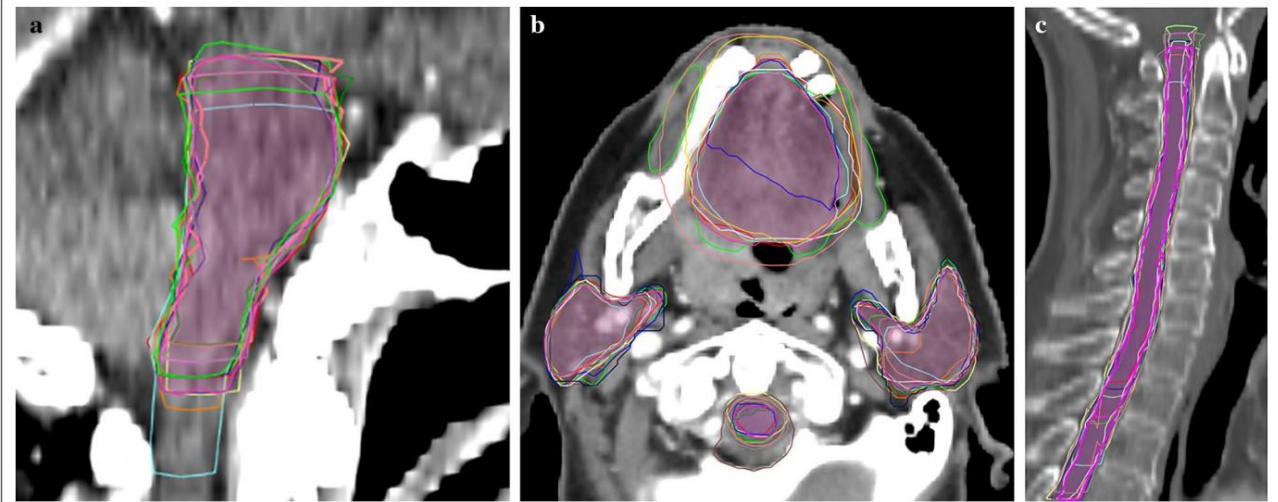
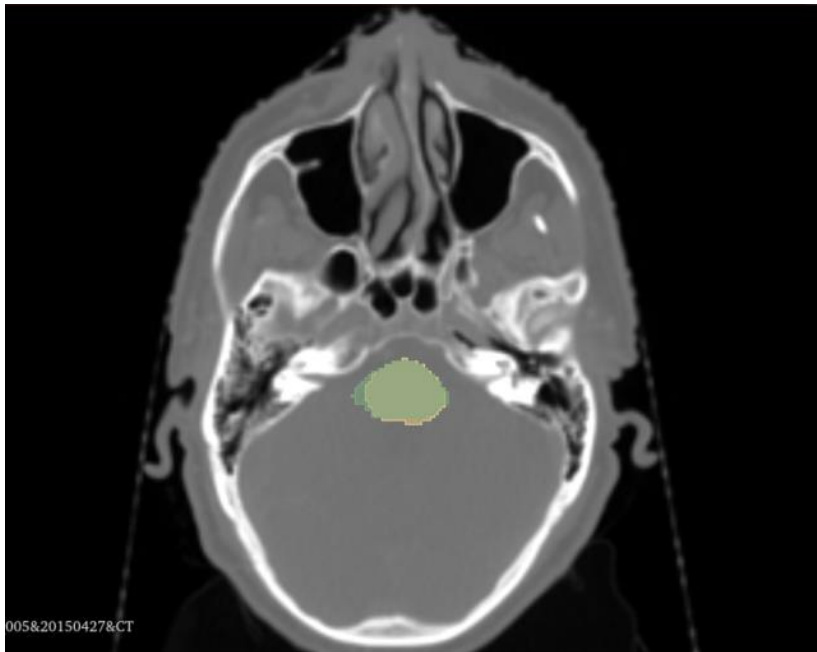
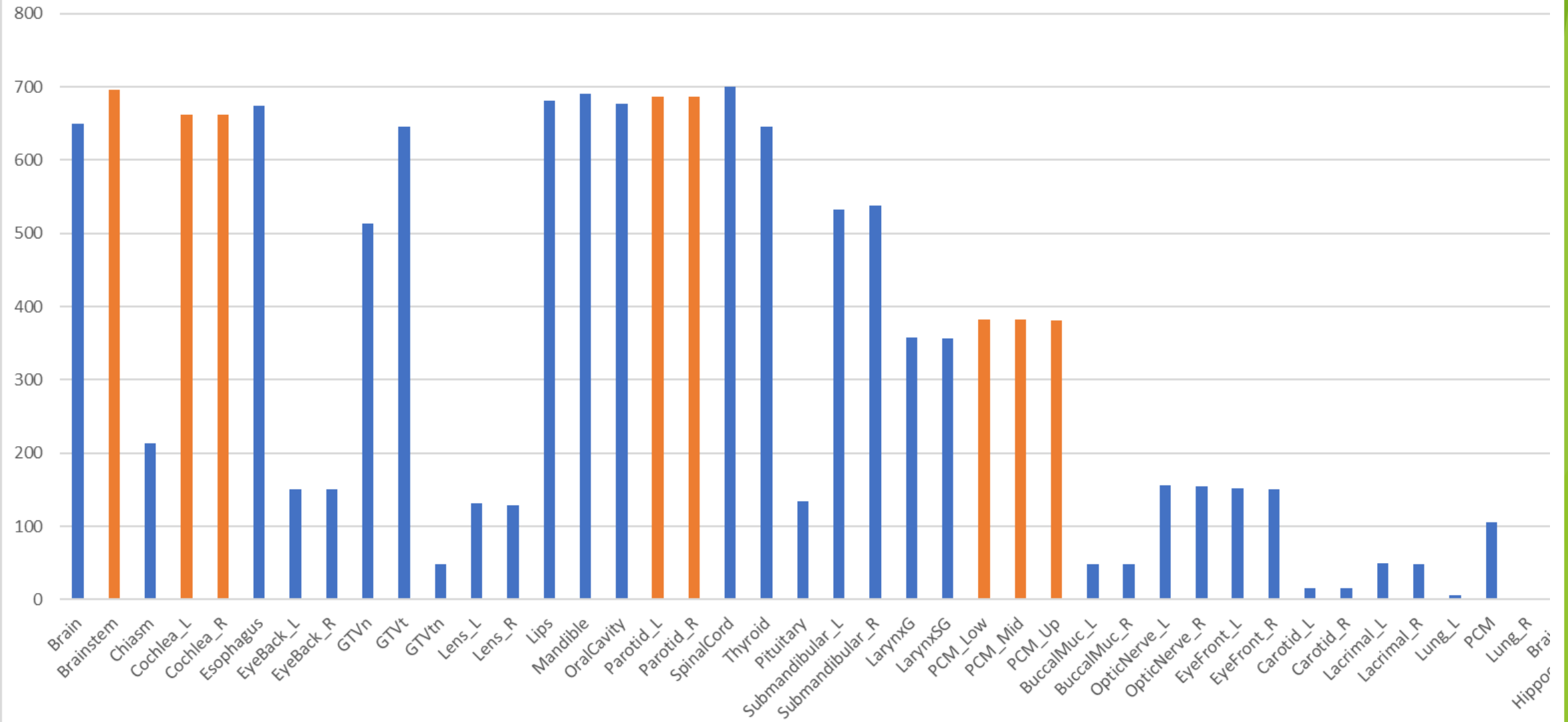


Fig. 3 CT images showing different OAR contours Reference contours according to the ICG (lilac) vs delineations from the different RO. **a** brainstem (sagittal plane): difference in cranial and caudal borders; **b** oral cavity, spinal cord and PG (axial plane): Inclusion of buccal mucosa (green contour) and teeth (orange and pink contours) by some RO. Variation in spinal cord and PG contours; **c** spinal cord (sagittal plane): difference in cranial and caudal borders. ICG international consensus guidelines, PG parotid glands, RO radiation oncologist

Indtegninger af OARs for 713 patienter



What's next?

Short run

- ▶ Data processing/cleaning
 - ▶ E.g designing classes for data handling
- ▶ Understanding volume comparison measures
 - ▶ DICE, Hausdorff, Mean Surface Distance, Added Path Length

Long run

- ▶ Design benchmarking tools
- ▶ Benchmarking performance
- ▶ Visualization of results
 - ▶ Similarity plots
 - ▶ Organ delineation heatmaps
- ▶ Writing a scientific paper