

# Performance Testing of Auto-Segmentation Algorithms

Alex Kolby, Eskild H. Andersen, Simon Rydder

March 2022

## 1 Hvad går projektet overordnet ud på?

Inden strålebehandling af kræftpatienter indtegner læger risikoorganer på computertomografi skanninger (CT-scans). Denne manuelle indtegningsproces er tidskrævende og præget af usikkerheder, og derfor ønskes det at sammenligne performance af forskellige auto-segmentationsalgoritmer til indtegningsproces af risikoorganer.

Sammenligningen skal herfra ske på baggrund af udregning af forskellige metrikker, der er et mål for graden af lighed mellem segmentationer.

## 2 Hvad er den faglige problemstilling?

I praksis benyttes en auto-segmentationsmetode kaldet ATLAS, der tager udgangspunkt i en placering af organer for et gennemsnitligt menneske, til at mindske tidsforbruget på indtegningsproces for lægerne. Nyere metoder, såsom Deep Learning, er også begyndt at blive brugt i udlandet, som viser sig lovende, men bliver ikke brugt klinisk i Danmark endnu. Det er disse forskellige segmentationsalgoritmer vi skal benchmarke mod hinanden, for at se om segmentationsalgoritmerne reelt set er tidsbesparende for lægerne.

Der findes forskellige metrikker til at sammenligne segmentationer (eller billeder), såsom DICE, Hausdorff og 'Added Path Length'. DICE er en metrik der måler snittet mellem de to segmenter, men er stærkt påvirket af segmenternes størrelse. Hausdorff kigger på afstanden mellem kanterne på de to segmenter. Sidstnævnte er 'Added Path Length'/'Edited Path Length', som har vist sig at være bedst korreleret med lægens tidsforbrug, som måler længden af ikke overlappende kant. Grundet forskelligheden er sammenligning af resultaterne for metrikkerne relevant.

### 3 Hvilke data kigger I på, og hvad skal man være særlig opmærksom på?

Datasættet består af semi 3D CT-scans og dertilhørende segmenteringer af organer. For hver patient er der 3 forskellige segmenteringsmetoder samt en "ground truth" segmentering. Denne er en manuel indtegnning af lægen, og dermed det tætteste vi kan komme på den sande segmentering. De 2 første segmenteringsmetoder er Deep Learning baseret og ikke anvendt i praksis endnu. Den sidste er den segmenteringsmetode der bliver anvendt i dag, kendt som ATLAS.

Da mange af patienterne er blevet brugt til at træne Deep Learning Modellerne, har vi "kun" 73 patienter i testsættet, hvor der er lidt variation i hvilke risiko organer der er indtegnet.

### 4 Hvad håber I at finde ud af?

Vi håber at finde ud af, at Deep Learning segmentationsalgoritmen performer bedre end ATLAS metoden, så lægerne kan få et bedre værktøj til at hjælpe med segmentationen af organer, når folk skal behandles for kræft.

### 5 Er der særlige udfordringer?

En af de større udfordringer var opgaven at implementere en nyere performance metrik, kendt som "Added path length"/"Edited path length", i python. Efter som der generelt var begrænset viden om denne metrik i forvejen.

En anden udfordring var at sætte sig ind i det lægefaglige stof samt hele forståelsen for forskellige måder at beregne ligheden mellem to sæt af billeder.

### 6 Hvad skal projektet munde ud i?

Projektet skal munde ud i form for visuel præsentation af resultaterne, der forhåbentligt kan overbevise de lægefaglige kyndige på Aarhus Universitet, om at Deep Learning metoden outperformer ATLAS.