

EXAMENSARBETE Vision transformers for segmenting organs and tissues in CT scans of arbitrary imaging ranges**STUDENTER** Morris Thånell, Petter Melander**HANDLEDARE** Ida Arvidsson (LTH), Filip Winzell (LTH), Johan Brynolfsson (Exini), Hannicka Sahlstedt (Exini)**EXAMINATOR** Kalle Åström (LTH)

Automatisk segmentering av organ i datortomografiska bilder oavsett bildsnitt

POPULÄRVETENSKAPLIG SAMMANFATTNING **Morris Thånell, Petter Melander**

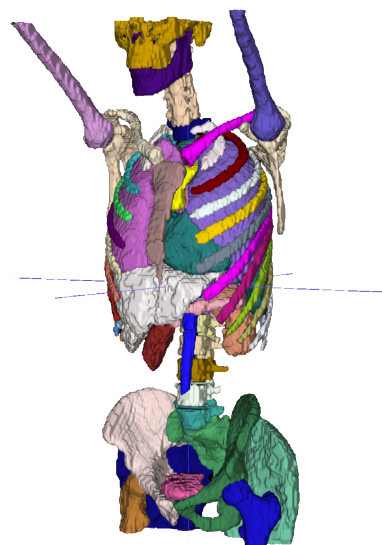
Automatisk segmentering av organ och vävnader i datortomografiska bilder kan hjälpa läkare att få anatomisk kontext inför planering av exempelvis kirurgi eller dosimetri. Vi har utvecklat två AI-modeller som kan göra just detta och göra det oavsett hur stor del av kroppen bilden är tagen på.

Segmentering innebär att man ger en klass till varje pixel i en bild för att markera ut olika objekt i bilden. I medicinska bilder kan dessa klasser till exempel vara lever, hjärta eller kranium. Våra modeller kan segmentera organ och vävnader i datortomografiska bilder (DT-bilder). DT är tredimensionell röntgen, vilken innebär att bilden är en 3D-volym som man kan se hela kroppen i, inte bara ett skikt som i vanlig röntgen. Det man kan se i DT-bilder är densiteten i alla delar av kroppen, vilket innebär att man till exempel ser skelettet tydligt eftersom det har hög densitet, och lungor tydligt för de har låg densitet. Eftersom man inte vill utsätta patienterna för onödigt mycket strålning tar läkare inte alltid bilder av hela kroppen utan ibland bara av mindre snitt. Vi har utvecklat modeller som kan segmentera ut organ och ben oavsett vilket snitt av kroppen som DT-bilden föreställer

Vi har använt oss av en form av AI som kallas deep learning. Deep learning-modeller är stora modeller som består av många lager, där varje lager löser en liten del av problemet.

I verkligheten är varje lager någon sorts matematisk funktion och vi har använt oss av en sort som

kallas transformer. Transformers utvecklades från början till textbaserade problem, där de tillåter en modell att förstå hur innebörden av ett ord påverkas av de andra orden i texten. Transformers är teknologin som möjliggör chat-bottar så som ChatGPT, där T:et står för transformer.



Segmentering visad i 3D.

Vi har använt oss av vision transformers som på ett liknande sätt tillåter modellen att förstå hur olika delar av DT-bilden hänger ihop.



DT-bild och segmentering visad i 2D.

Vi har använt två modeller, en färdig modell som heter Swin UNETR och en som vi har skapat själva som vi kallar vår slice-modell. Båda använder transformerlager för att se hur olika stycken av bilden hänger ihop. Swin UNETR delar upp den tredimensionella bilden i kubformade stycken

medan slice-modellen delar upp bilden i horisontella skikt och modellerna kollar hur dessa stycken hänger ihop.

Vi har tränat modellerna att segmentera 20 centimeter tjocka snitt av kroppen. När modellen ska segmentera större bilder, delar vi in bilden i 20 centimeter tjocka delar som vi segmenterar var för sig. Vi sammanfogar sedan de mindre delarna för att få ut en segmentering av hela bilden. När man segmenterar mindre snitt av kroppen är det svårt att veta exakt vilken ryggradskota och vilket revben som är vilket. Vi valde därför att inte numrera kotor och revben med våra modeller.

Vi kunde se att Swin-modellen fungerade lite bättre än slice-modellen, men båda fungerade bra. Vi är väldigt nöjda med att vi kunde göra segmenteringen av olika snitt av kroppen med en enda modell. Modellerna är lätta och snabba att köra, vilket gör att vi tror att de hade kunnat användas på riktigt i en sjukhusmiljö.