SAGIVTECH IMVC 2024

Improving robustness of large structures segmentation using partial annotations

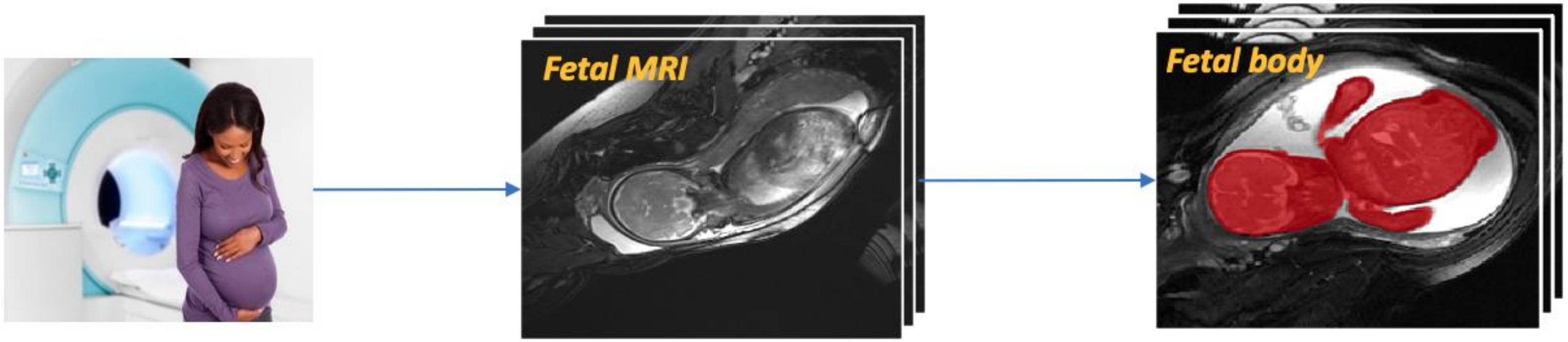
Dr. Bella Specktor Fadida,
Haifa University

Bella Specktor Fadida, Daphna Link Sourani, Liat Ben Sira, Elka Miller,
Dafna Ben Bashat, Leo Joskowicz

*This work is part of a PhD thesis at the Hebrew University of Jerusalem

Partial annotations for segmentation of large structures Introduction

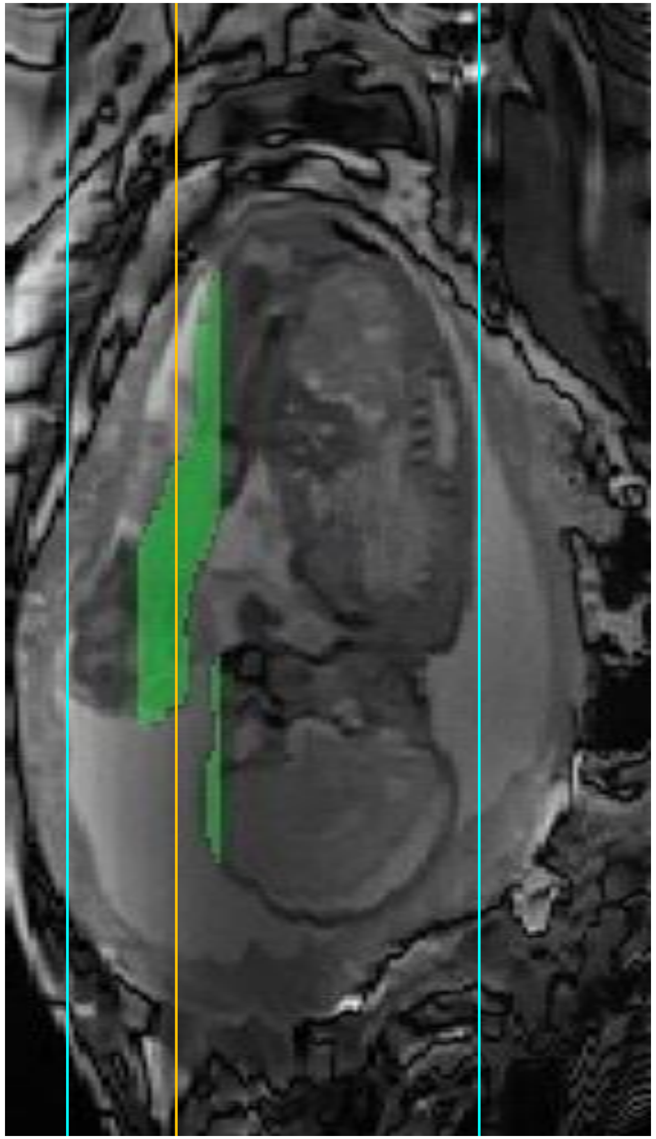
- Deep learning segmentation methods require large annotated datasets, whose manual segmentation is time-consuming and can take more than an hour for large structures
- Under low data regime, one can create more partialy annotated cases compared to fully annotated cases



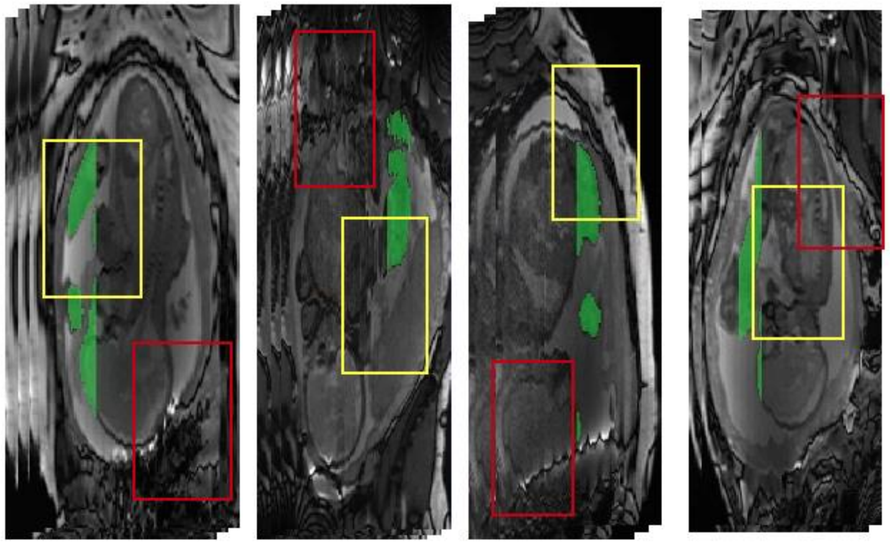
Method: Manual Partial Delineations

The user partially annotates scans with the algorithm guidance:

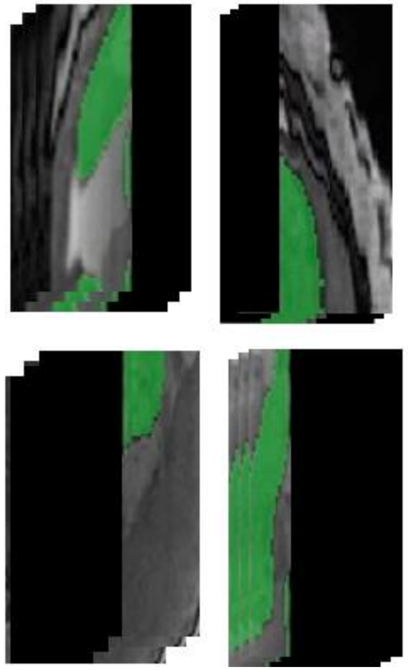
1. The uppermost and lowermost slices of the organ are manually selected by the annotator (turquoise line).
2. The algorithm randomly chooses a slice within the structure of interest (yellow line).
3. Consecutive slices are selected. The number of slices is determined by the chosen annotation percentage (green annotations).



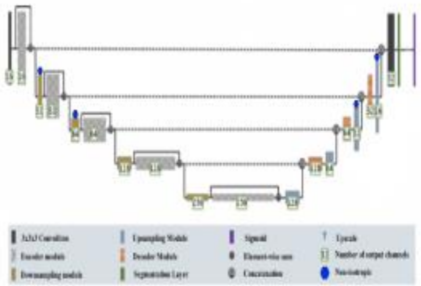
Method: Training with partial annotations



(1) Training input: saggital view of partially annotated scanns



(2) A batch of non-empty patches



(3) Training with a selective batch loss

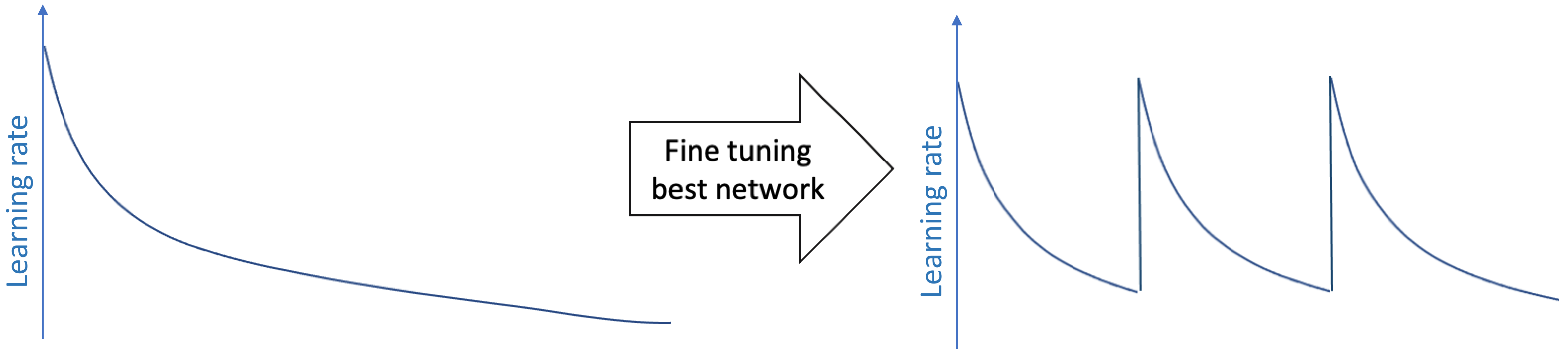
Method: Selective Dice loss

Let $\mathrm{T}' \subset \mathrm{T}$ and $\mathrm{R}' \subset \mathrm{R}$ be the ground truth in the annotated slices and the network result in the annotated slices, with minibatch voxels $t_i{}' \in T'$ and $r_i{}' \in R'$ respectively.

Selective batch Dice Loss ($\mathrm{L_{CD}}$) = $-\frac{2\sum_{N'} t_i{}' r_i{}'}{\sum_{N'} t_i{}' + \sum_{N'} r_i{}'}$

- Border slices are used by the loss function – considered "annotated slices"
- Large batch size of 8.
- Adding a binary mask specifying the locations of the annotated slices

Method: Two-step Training



1. Reduce on plateau scheme

2. Reduce on plateau with learning rate restarts scheme

Data and Experimental Design

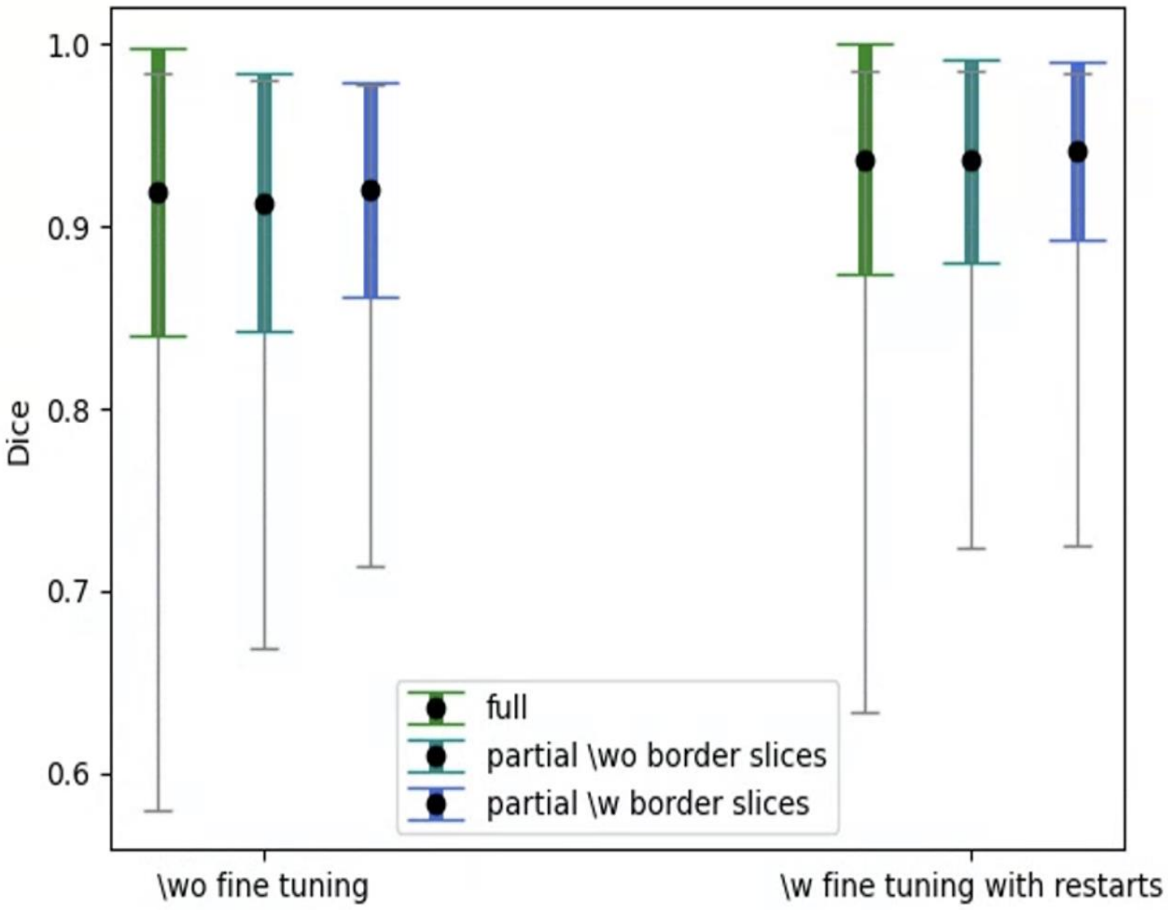
Data

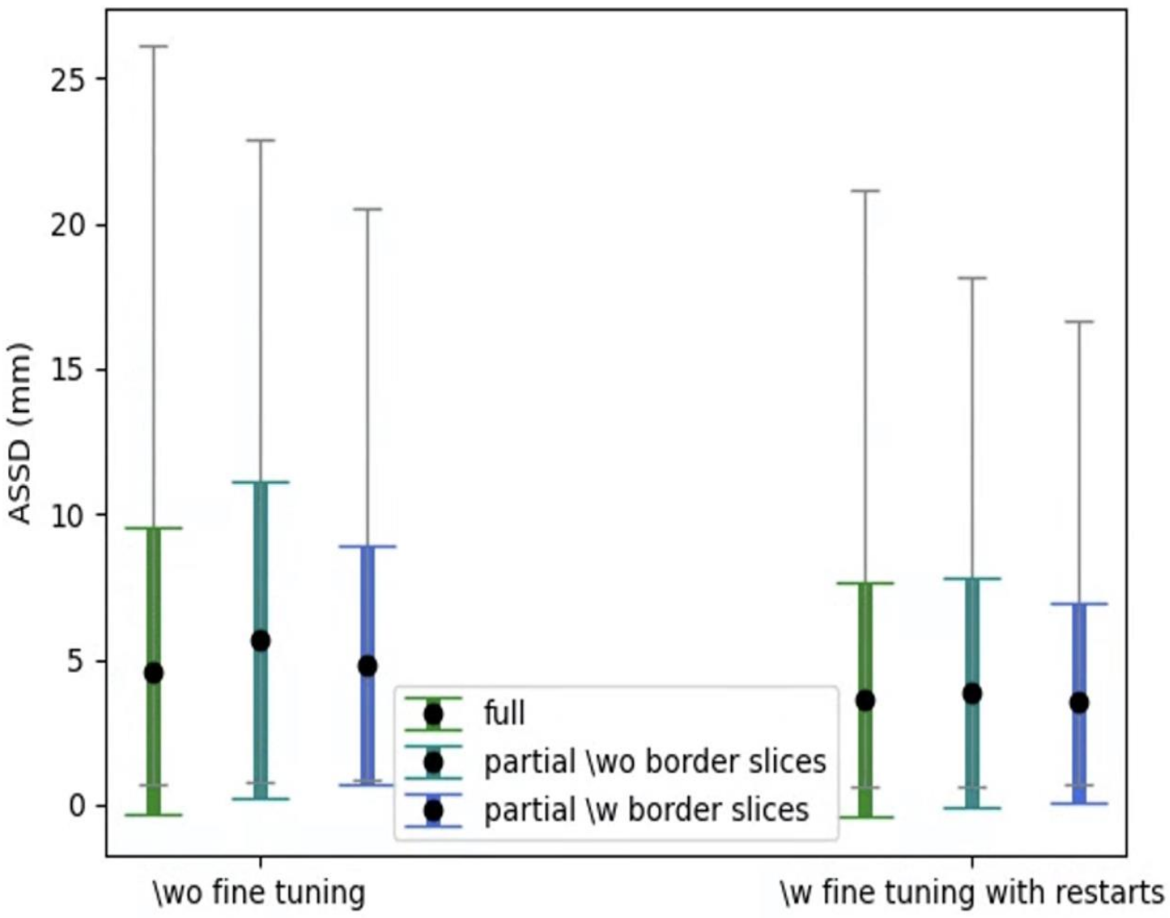
1. TRUFI body: 101 cases in total with gestational Age (GA) 28-39, 58 in-distribution (ID) test cases
2. FIESTA body: 137 cases in total. ID cases with GA 28-39 similar to training set (68 test cases) and most Out of Distribution (OOD) cases with GA 16-24 (33 test cases)

Experimental Design

- Training regime with 30 partially annotated cases and 20% annotated slices are compared to training with 6 fully annotated cases.
- The 6 cases are randomly chosen from the 30 partially annotated cases.
- Results are an average of 4 different randomizations.

Results on TRUFI body





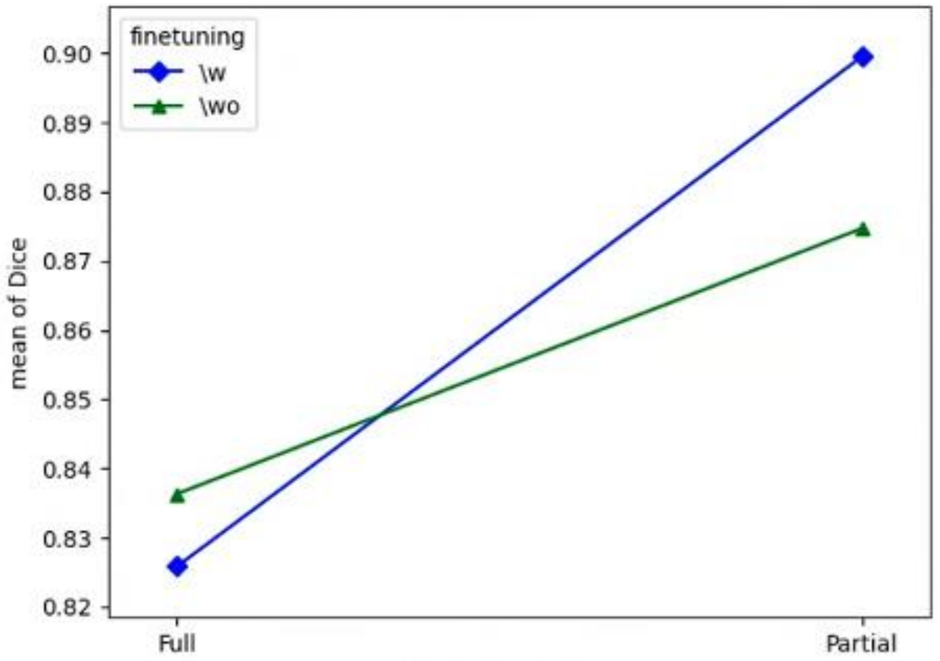
Results on FIESTA Body

Data distribution	Network training	Dice	Hausdorff (mm)	2D ASSD (mm)
In-Distribution (ID)	Full	0.959±0.044	34.51±37.26	2.15±2.33
	Full fine-tuned	0.964±0.040	32.98±36.86	1.88±2.07
	Partial	0.959±0.034	34.15±35.96	2.21±1.67
	Partial fine-tuned	0.965±**0.029**	31.89±35.82	1.90±**1.39**
Out-of-Distribution (OOD)	Full	0.836±0.178	39.34±29.26	7.46±10.61
	Full fine-tuned	*0.826±0.214*	*39.61±32.66*	*8.86±16.54*
	Partial	0.875±0.091	36.19±21.44	5.47±3.92
	Partial fine-tuned	**0.899±0.067**	**30.37±18.86**	**4.00±2.26**

Statistical Analysis for OOD Data

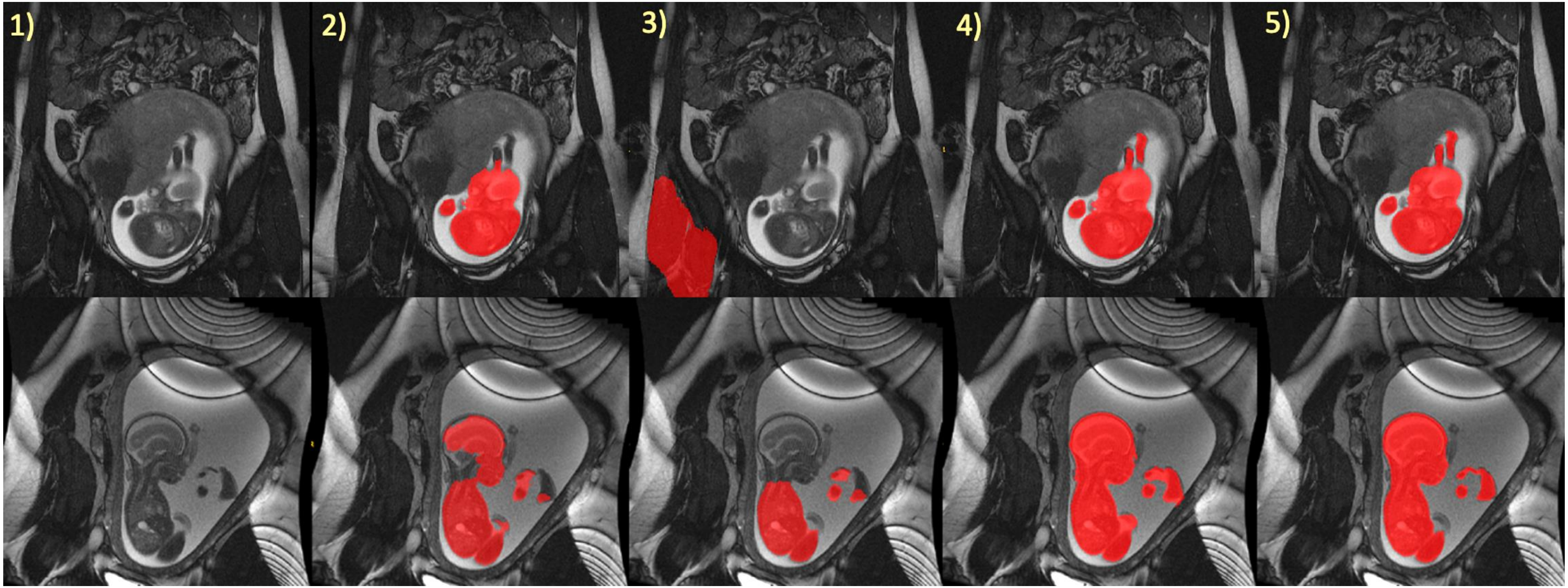
Annotation strategy	Dice	Hausdorff Distance	2D ASSD
Fine-tuning (w/wo)	F=1.69 p=0.202	F=6.15 p=0.019*	F=0.007 p=0.934
Annotation strategy (Full / Partial)	F=8.96 p=0.005**	F=5.83 p=0.022*	F=6.473 p=0.016*
Interaction	F=7.74 p=0.009**	F=9.88 p=0.004**	F=7.78 p=0.009**

Repeated measurements two-way ANOVA. Significance codes: *<0.05; **<0.01.



Interaction plot for Dice score.

Results on FIESTA Body



Full annotations no fine-tuning

Full annotations with fine-tuning

Partial annotations with fine-tuning

Ground Truth

Conclusions

- We have presented a new method for using partial annotations for large structures.
- The method demonstrated better robustness in a low data regime compared to full annotations.
- We also presented a simple two-step optimization scheme for low data regime that combines fine-tuning with learning rate restart.
- The optimization was useful for partial annotations regime on both ID and OOD data. For full annotations it decreased performance on OOD data.