

EP04.06

Evaluation of synthetic-data-trained foundation model for accelerating annotation in gynecological ultrasound

Fruzsina Dvorzsák¹, Alinka Olajos-Horváth¹, Mária Prosszer¹, Vanda Czipczer¹
¹GE Healthcare Magyarország Kft., Budapest, Hungary

Objectives

Foundation models such as Segment Anything Model (SAM) are transforming image annotation workflows, yet their performance on medical images remains limited by the lack of robust, well-annotated datasets. These issues were recently addressed by finetuning SAM2 on fully synthetic 3D data replicating anatomy and tissue boundaries. We aimed to evaluate this new model, SynthFM, on gynecological ultrasound.

Methods

3D Slicer was extended with a module integrating foundation models, operated with point prompts and bounding box. Five ultrasound volumes were segmented for uterus and endometrium by two observers. SynthFM segmentations were compared to previous SAM2 results and expert-validated ground truth annotations (Figure 1). Annotation times were recorded.

Results

SynthFM reduces average operation time by 65% for uterus and 61% for endometrium compared to SAM2 annotations (Table 1). Model performance was evaluated with Dice Similarity Coefficient (DSC) and 95th-percentile Hausdorff Distance (HSD95). SynthFM shows increased accuracy and fewer large boundary errors compared to SAM2 for both uterus (DSC=0.91) and endometrium (DSC=0.78).

Organ	Average segmentation times (h:mm)			Average accuracy (DSC)			Average greatest boundary error (HSD95)		
	Manual*	SAM2*	SynthFM	Manual*	SAM2*	SynthFM	Manual*	SAM2*	SynthFM
uterus	2:22	0:16	0:05 (Δ-65%)	0.912	0.861	0.909	8.623	14.812	6.915
endometrium	0:29	0:11	0:04 (Δ-61%)	0.775	0.627	0.784	4.963	10.618	5.121

*based on recently published results

Table 1.
Model operation times and quality of segmentations.

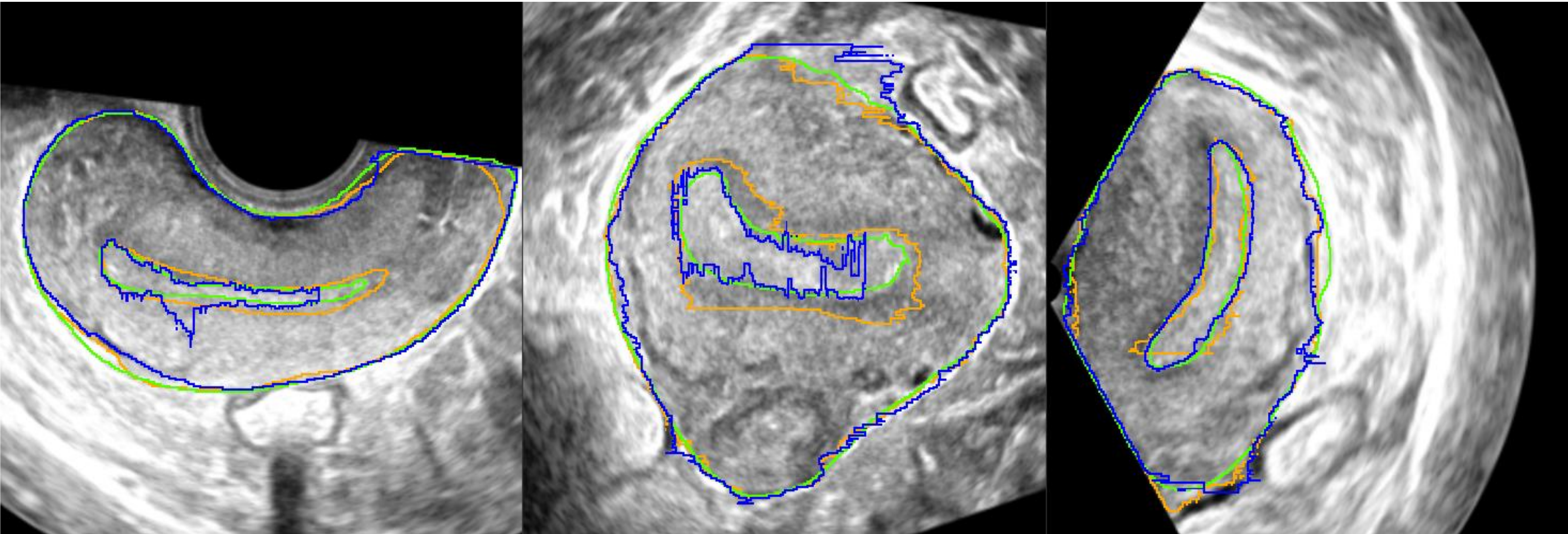


Figure 1.
Example of SAM2 (blue), SynthFM (orange) and ground truth (green) segmentations on three orthogonal planes.

Conclusions

Synthetic data finetuning helps in annotation acceleration with precise volumetric auto-segmentation. Although we observed a clear performance increase compared to SAM2, a limitation is that the previous study evaluated a SAM2 variant with different configurations, indicating that further research is needed for fully comparable results.

Acknowledgements

The project has been implemented with the support provided by the Ministry of Culture and Innovation of Hungary from the National Research, Development and Innovation Fund, financed under the **2024-1.1.2-NAGYVÁLL_FÓKUSZ-2025-00006** fundingscheme.