Automated Segmentation of Pelvic/Ovarian and Omental Lesions on CT Images of Ovarian Cancer Patients using Deep Learning

Thomas Buddenkotte[1,2], Leonardo Rundo[2,3], Ramona Woitek[2,3,4], ...
Evis Sala[2,3], Carola-Bibiane Schönlieb[1]

[1]Department of Applied Mathematics and Theoretical Physics, University of Cambridge, [2]Department of Radiology, University of Cambridge, [3]Cancer Research UK Cambridge Centre, University of Cambridge, [4]Department of Biomedical Imaging and Image-guided Therapy, Medical University Vienna

UNIVERSITY OF CAMBRIDGE

CCIMI
CANTAB CAPITAL INSTITUTE FOR THE MATHEMATICS OF INFORMATION

Abbreviations

High grade serous ovarian carcinoma (HGSOC), neoadjuvant chemotherapy (NACT), computed tomography (CT), Dice similarity coefficient (DSC), no-new-UNet (nnU-Net), state-of-the-art (SOTA).

Introduction

- HGSOC patients show no increase in overall survival since 20 years.
- Currently used RECIST measurements show limited correlation with patient outcome.
- Volumetric tumor assesment and radiomics show more promising performance.
- Goal: automated segmentation of the main lesion sites.
- Objectives:
  - How does the model performance change when the amount of training data is increased?
  - Despite large differences in tumor extent, can one deep learning model segment both pre- and post-NACT scans?
  - How does the model perform when compared to unrevised segmentations provided by trainee radiologists?

Materials and Methods

- Three distinct datasets of HGSOC patients provided with gold standard manual segmentations.
- 15 patients from Dataset # 2 with additional unrevised segmentations performed by a trainee radiologist.
- Use the two-stage cascade of nnU-Net as a deep learning model.
- Target only the two most common lesion sites: pelvic/ovarian and omental lesions.

Table: Composition of the three analyzed ovarian cancer CT datasets. The pixel spacing and slice thickness are displayed as median (minimum-maximum). For both lesion types, the scans are displayed as number of scans containing the lesion (average gold standard volume, average gold standard number of conntected components).

	Dataset #1	Dataset #2	Dataset #3
Pixel spacing	0.68 (0.53-0.93)	0.76 (0.61-0.96)	0.77 (0.57-0.98)
Slice thickness	5.0 (1.25-5.0)	5.0 (1.5-5.0)	5.0 (2.0-7.5)
Pre-treament scans	157	53	71
pelvis/ovaries	144 (346.4cm^3, 2.6)	53 (322.2cm^3, 3.2)	69 (380.7cm^3, 2.4)
omentum	120 (141.3cm^3, 8.6)	52 (202.3cm^3, 6.7)	56 (146cm^3, 5.7)
Post-treament scans	119	51	N/A
pelvis/ovaries	102 (175.0cm^3, 2.0)	49 (154.1cm^3, 1.9)	N/A
omentum	78 (84.47cm^3, 3.9)	46 (56.6cm^3, 3.7)	N/A

Experiments

- Use four different configurations of training and test sets to address the main questions.
- Evaluate the performance in terms of mean DSC.
- Repeat for both lesions sites independently.

Table: Training and test configurations used in the evaluation.

Experiment	A	B	C	D
Training datasets	#1	#1	#1, #2	#1, #3
Scan types	Only pre-treatment	All	All	All
Test datasets	#2, #3	#2, #3	#3	#2

Results

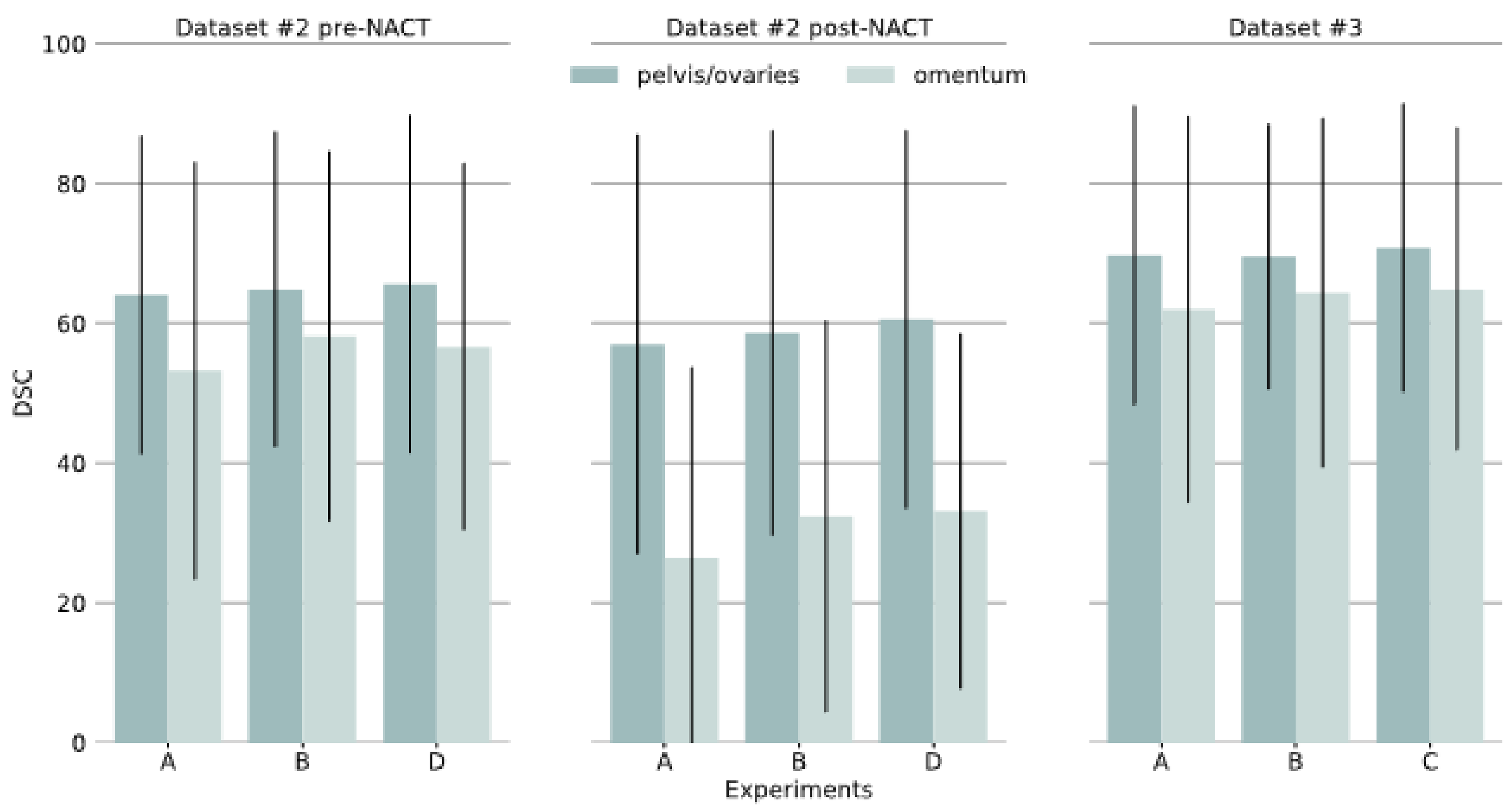

Figure: DSC on the unseen external data for the four experiments A, B, C, and D. The scores for the pelvic/ovarian and omental lesions are shown in dark and light green, respectively.

- Increasing the training set size improved the mean DSC except for two cases.
- Large standard deviation.
- Correlation of lesion volume with DSC (data not shown).

Table: Evaluation of trainee *versus* automated performance on Dataset #2. The p-values were computed using a Wilcoxon test to compare the DSC values.

		Pre-treatment		Post-NACT	
Metric	Segmentation	Pelvic/ovarian	Omentum	Pelvic/ovarian	Omentum
DSC	trainee	70.51	63.47	71.41	40.45
	automated	71.96	51.67	66.42	28.76
p-values		0.421	0.064	0.358	0.173

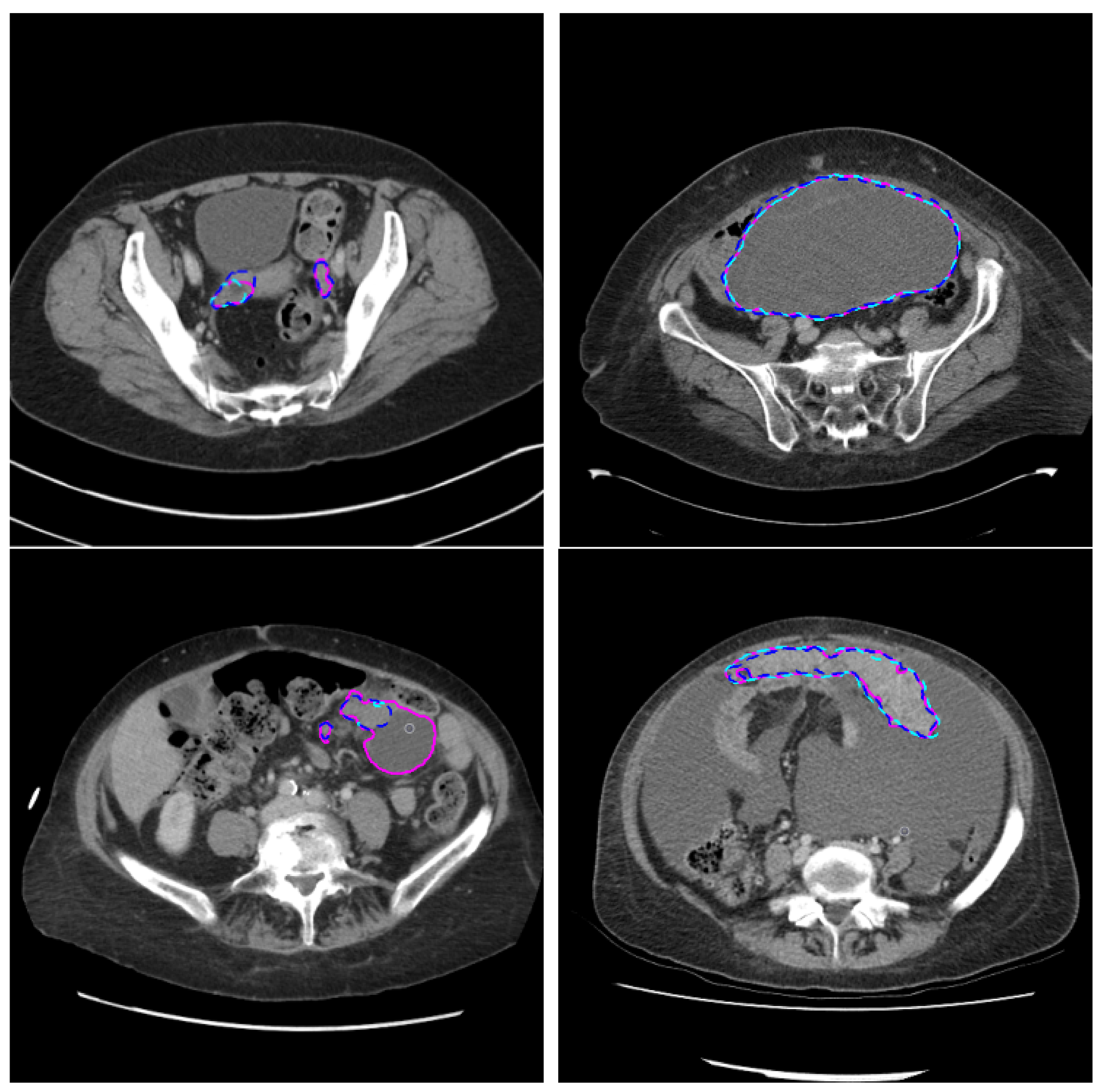

Figure: Segmentation examples of automated, gold standard (GS) and trainee segmentations (TA) denoted with blue, magenta and cyan lines, respectively.

Conclusions

- Automated segmentation of HGSOC is feasible.
- The mean test DSC does not always increase with training set size.
- Automated methods can segment both pre- and post-NACT treatment scans with a single model.
- SOTA deep learning models perform not significantly different from trainee radiologists.