JULY 12–16 2020 VIRTUAL
JOINT AAPM | COMP MEETING
EASTERN TIME [GMT-4]

Cascading Deep Multi-Label Network for CT Liver and Spleen structure Segmentation: Learning from imperfect clinical data

R. Haq[1], A. Jackson[1], A. Apte[1], M. Montovano[2], A. Wu[2], J.O. Deasy[1], and M. Thor[1]

[1]Department of Medical Physics, Memorial Sloan Kettering Cancer Center, New York, US
[2]Department of Radiation Oncology, Memorial Sloan Kettering Cancer Center, New York, US

Memorial Sloan Kettering Cancer Center

INTRODUCTION

- Radiation-induced decrease in lymphocytes as a result of spleen and liver irradiation during treatment planning has been linked to immunosuppression, resulting in compromised patient recovery [1, 2].
- Performing clinical outcomes analyses requires robust and repeatable contours of the liver and spleen.
- However, the internal clinical data available for auto-segmentation consisted of well-curated spleen contours but unverified quality of the liver contours.

AIM

Provide an open-source tool to automatically segment immune-related organs-at-risk regardless of clinical contour quality.

To facilitate this, we built and validated a Cascading Deep Learning multi-label Segmentation (cDLS) framework to automatically segment liver and spleen contours.

METHOD

- 172 internal Computed Tomography (CT) scans with well-curated spleen contours and unverified liver clinical contours
  - Patients treated with radiation therapy for esophageal cancer.
- Leveraged transfer learning to mitigate the effect of learning from unverified contours.
  - The cDLS framework learned contextual information from 20 external CT scans from the CHAOS publicly available dataset [3] consisting of well-curated liver contours.

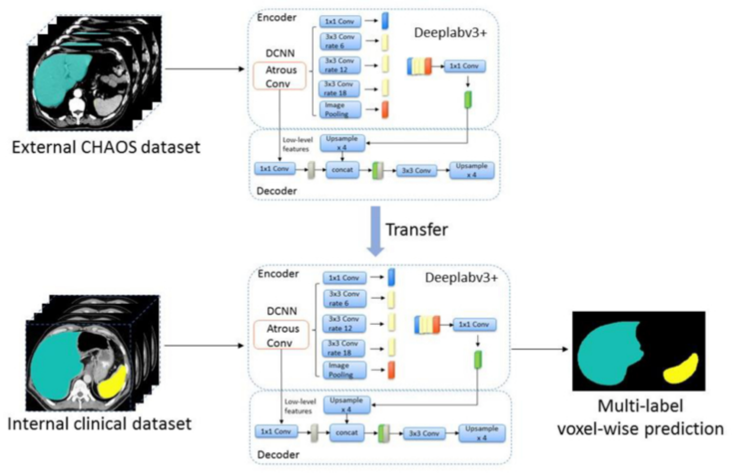

Figure 1: Overview of our proposed method. The developed cDLS model utilized the Deeplabv3+ architecture with a Restnet101 backbone. Learned representations from the external dataset were transferred as pretrained weights for training the internal dataset for multi-label voxel-wise segmentation of the liver and the spleen.

cDLS Model	Internal dataset	External Dataset
Training/Evaluation data	132/20 CT scans	17/3 CT scans
Hold-out test data	20 CT scans	none
Contours	Spleen (well-curated) Liver (unverified)	Liver (well-curated)
Data Augmentation	Random cropping, horizontal and vertical flipping, rotation by 10 degrees	
Data preprocessing	Normalization (0-255), resizing (512 x 512)	
Training Parameters	*Backbone:* Resnet101, *learning rate:* 0.01, *batch size:* 8, *Epochs:* 25 (external dataset) and 50 (internal dataset), *training loss:* cross-entropy, *output stride:* 16, *scheduler:* "policy" learning rate.	

Evaluation Metrics

- Dice Similarity Metric (DSC)
- 95th Percentile of Hausdorff Distance (HD95) (mm)
- Mean Dose Volume Histogram (DVH) comparison.
  - Wilcoxon rank-sum test applied for statistical significance.

RESULTS & DISCUSSION

Quantitative Evaluation:

Figure 2 displays the cDLS contours achieved DSC accuracies of

- Liver = 0.95(0.93-0.96),
- Spleen = 0.94(0.91-0.94), and

HD95, a measure of the 95th percentile of the worst surface error, of

- Liver =4.9mm(4.2mm-5.8mm),
- Spleen = 3.3mm(2.9mm-3.7mm).

The model reduced about half an hour of manual contouring to less than 10 seconds of auto-segmentation time.

No statistically significant difference was found between the DVH metric comparison extracted from the reference against the cDLS contours. (**Table 1**).

Qualitative evaluation:

The automatically generated contours were able to follow anatomical boundary for both the liver and the spleen organs with fidelity (**Figure 3**).

Comparison between the unverified reference liver contours and automated cDLS contours (**Figure 4**) for some patients showed that the reference liver contours sometimes contained the Inferior Vena Cava (IVC) structure boundary, which the cDLS contours were able to successfully exclude.

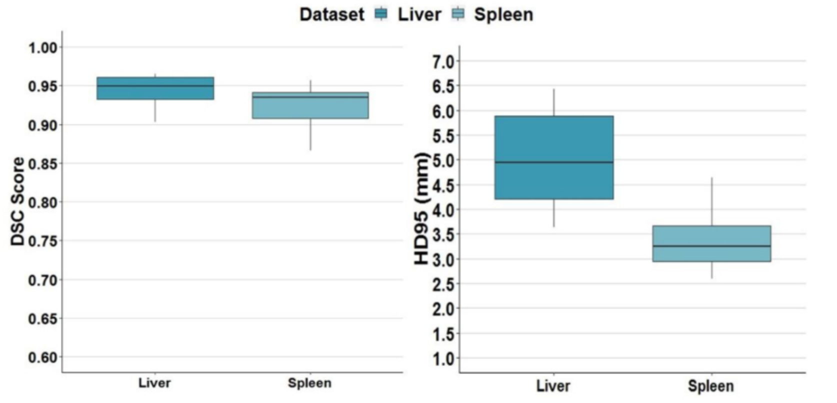

Figure 2: Quantitative Evaluation: (a) Median Dice Similarity Coefficient (DSC) and (b) 95th Percentile of Hausdoff Distance (HD95) (mm) for 20 abdominal radiation therapy CT scans comparing auto-generated contours using cDLS architecture against manually delineated expert contours

Structure (metric)	cDLS to Expert contour difference (%)	p-value*
Liver Mean Dose (Gy)	0.91 (-1.68 – 3.49)	0.904
Spleen Mean Dose (Gy)	0.24 (-4.25 – 3.76)	0.980

Table 1: Quantitative Mean Dose (Gy) evaluation of auto-generated segmentations of the liver and the spleen against reference contours for 20 CT scans of patients previously treated with radiotherapy in the clinic. *Statistical comparison between the dose difference was performed using the Wilcoxon signed rank-sum test. Percentage difference was calculated as Difference (%) = $\left(\frac{DLS_{volume} - Expert_{volume}}{Expert_{volume}}\right) \times 100$

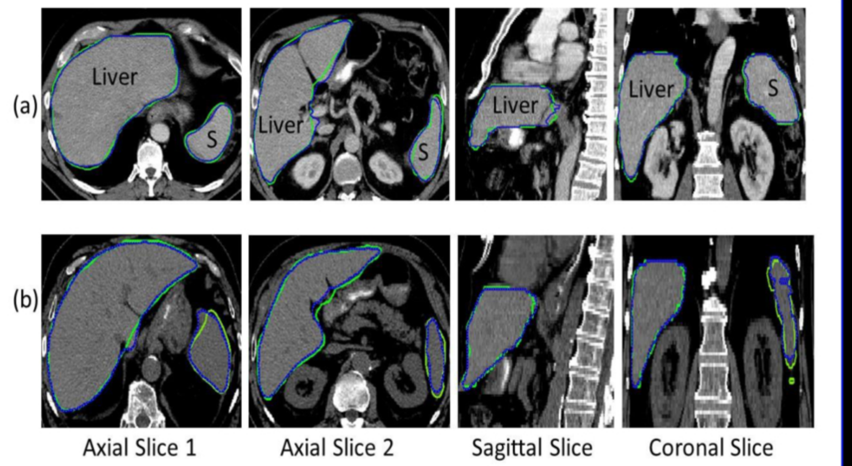

Figure 3: Comparison of auto-generated segmentations using developed cDLS architecture (blue) against reference manual clinical segmentation (green) for two randomly selected patients (a-b). S = Spleen.

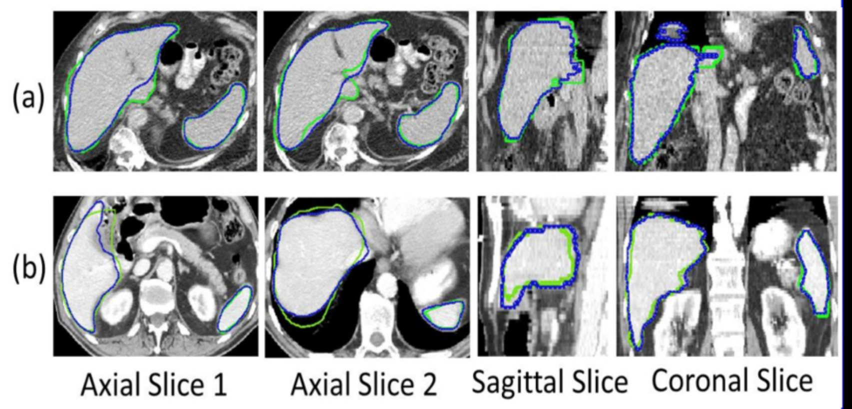

Figure 4: The highest disagreement observed between the auto-generated cDLS contours (blue) against the reference clinical contours (green) for two sample patients (a-b). The automated contours were able to successfully exclude IVC structure within the liver segmentation (patient (a)) and better capture the superior and inferior aspects of the anatomical structure (patient (b)).

CONCLUSIONS

- The developed model was able to learn representations from external, well-curated data resulting in robust and repeatable automatically generated liver and spleen segmentations.
- The accuracy was judged adequate for extracting dose-volume information for outcomes analyses.
- The cDLS model is distributed as part of CERR's[4] Model Implementations library.

REFERENCES

[1] Lui J, Zhao Q, *et al*. Radiation-related lymphopenia is associated with spleen irradiation dose during radiotherapy in patients with hepatocellular carcinoma. *Radiat Oncol* 2017;12:90.

[2] Venkatesulu BP, Sharma A, Pollard-Larkin JM, et al. Ultra high dose rate (35 Gy/sec) radiation does not spare the normal tissue in cardiac and splenic models of lymphopenia and gastrointestinal syndrome [published correction appears in Sci Rep. 2020 Jun 30;10(1):11018]. *Sci Rep.* 2019;9(1):17180. Published 2019 Nov 20. doi:10.1038/s41598-019-53562-y

[3] Kavur AE, Selver MA, et al. CHAOS – Combined (CT_MR) Healthy Abdominal Organ Segmentation Challenge Data. Zenodo, 11-April-2019.

[4] Apte A.P., Iyer A. et al. Library of deep-learning image segmentation and outcomes model-implementations. *Physica Medica.* 2020; 73: 190-196

ACKNOWLEDGEMENTS

This research is partially supported by NCI R01 CA198121.

CONTACT INFORMATION

Rabia Haq, Ph.D.

haqr@mskcc.org

https://www.linkedin.com/in/rabiahaq/