Clinical implementation of deep learning: Automatic contouring via U-Net architecture

Matthew Cooper[1]
Simon Biggs[2]
Yu Sun[1]
Matthew Sobolewski[2]

[1]Institute of Medical Physics, The University of Sydney
[2]Riverina Cancer Care Centre, Cancer Care Associates.

Thesis: github.com/matthewdeancooper/masters_thesis

Contouring - Current limitations

Variability

- Large intra and inter-observer variance (IOV).[1]
- AAPM TG275 risk assessment - multiple human-factor failure modes in RT.[2]

Time constraints

- Atlas methods $\Longrightarrow$ significant correction times.[3]

Deep learning potential

- Shown to reduce IOV and contouring time.[3]
- Significant improvement cf. atlas methods (time & accuracy).[4]

---

[1] Dale Roach et al. "Multi-observer contouring of male pelvic anatomy: Highly variable agreement across conventional and emerging structures of interest". In: *Journal of Medical Imaging and Radiation Oncology* 63.2 (2019), pp. 264–271. DOI: 10.1111/1754-9485.12844

[2] Eric Ford et al. "Strategies for effective physics plan and chart review in radiation therapy: Report of AAPM Task Group 275". In: *Medical Physics* 47.6 (2020), e236–e272. DOI: https://doi.org/10.1002/mp.14030

[3] Shalini K Vinod et al. "A review of interventions to reduce inter-observer variability in volume delineation in radiation oncology". In: *Journal of Medical Imaging and Radiation Oncology* 60.3 (2016), pp. 393–406. DOI: 10.1111/1754-9485.12462

[4] Stanislav Nikolov et al. *Deep learning to achieve clinically applicable segmentation of head and neck anatomy for radiotherapy*. 2018. arXiv: 1809.04430 [cs.CV]

Research goals

Model 1: QA tool (prostate cancer).

- Compare model and expert contours to identify macro contouring errors
- Patient, bladder, rectum volumes.

Model 2: Canine vacuum bag.

- Automate time consuming aspect of canine RT.
- Previously, manual vacuum bag contouring $\approx$ 30 min.

Goal: Performance similar to human experts.

- Performance metric (sDSC) that takes into account expert IOV.[4]
- Stronger correlation with correction time cf. DSC.[5]

[3] Shalini K Vinod et al. "A review of interventions to reduce inter-observer variability in volume delineation in radiation oncology". In: *Journal of Medical Imaging and Radiation Oncology* 60.3 (2016), pp. 393–406. DOI: 10.1111/1754-9485.12462

[4] Stanislav Nikolov et al. *Deep learning to achieve clinically applicable segmentation of head and neck anatomy for radiotherapy*. 2018. arXiv: 1809.04430 [cs.CV]

[5] Femke Vaassen et al. "Evaluation of measures for assessing time-saving of automatic organ-at-risk segmentation in radiotherapy". In: *Physics and Imaging in Radiation Oncology* 13 (2020), 1–6. ISSN: 2405-6316. DOI: 10.1016/j.phro.2019.12.001

Performance - Surface dice similarity coefficient (sDSC)

$$DSC_{1,2} = \frac{2|M_1 \cap M_2|}{|M_1| + |M_2|}$$

$$sDSC_{1,2}^{(\tau)} = \frac{|S_1 \cap B_2^{(\tau)}| + |S_2 \cap B_1^{(\tau)}|}{|S_1| + |S_2|}$$

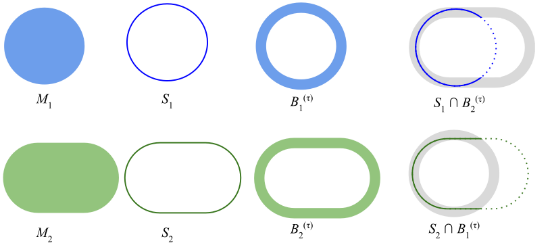


Figure: DSC is a volumetric overlap score, sDSC is a surface overlap score - the percentage of surface contoured within an organ specific tolerance representative of expert IOV.[3]

[4] Stanislav Nikolov et al. *Deep learning to achieve clinically applicable segmentation of head and neck anatomy for radiotherapy.* 2018. arXiv: 1809.04430 [cs.CV]

All happy models are alike...

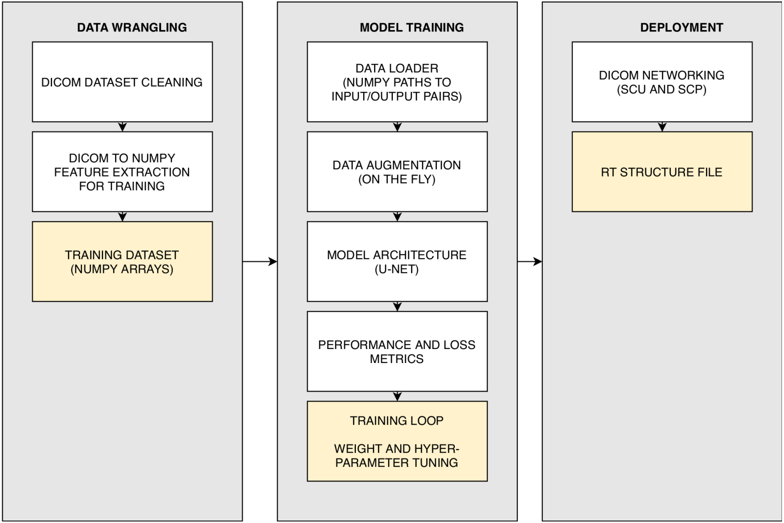


Figure: Modules required for end-to-end deep learning model deployment.

All happy models are alike...

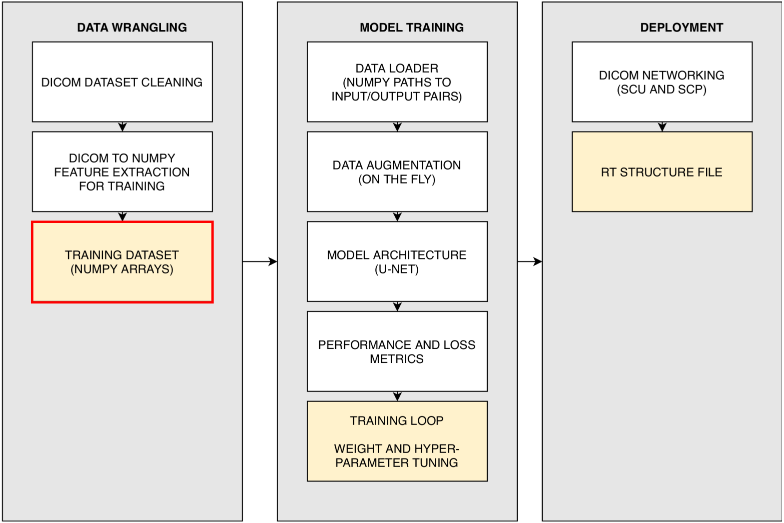


Figure: Modules required for end-to-end deep learning model deployment.

Method: Data augmentation

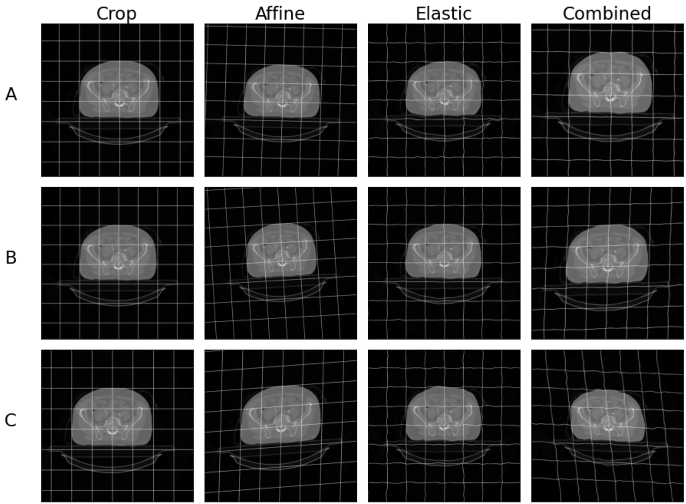


Figure: Data augmentation to improve model robustness.

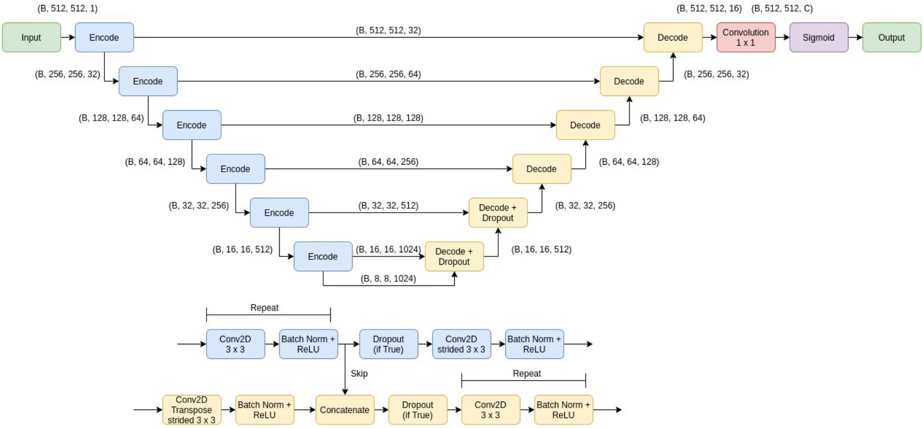


Figure: Modified 2D U-net architecture:[8] Composed of encoding (blue) and decoding blocks (yellow). MaxPooling layers replaced by strided convolution[11]. Added batch normalisation[12] and final sigmoid activation.[3] Tensor dimensions (Batch size, X, Y, Channels).

---

[3] Stanislav Nikolov et al. *Deep learning to achieve clinically applicable segmentation of head and neck anatomy for radiotherapy*. 2018. arXiv: 1809.04430 [cs.CV]

[8] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. *U-Net: Convolutional Networks for Biomedical Image Segmentation*. 2015. arXiv: 1505.04597 [cs.CV]

[11] Jost Tobias Springenberg et al. *Striving for Simplicity: The All Convolutional Net*. 2014. arXiv: 1412.6806 [cs.LG]

[12] Sergey Ioffe and Christian Szegedy. *Batch Normalization: Accelerating Deep Network Training by Reducing Internal Covariate Shift*. 2015. arXiv: 1502.03167 [cs.LG]

Method: Why downsample?

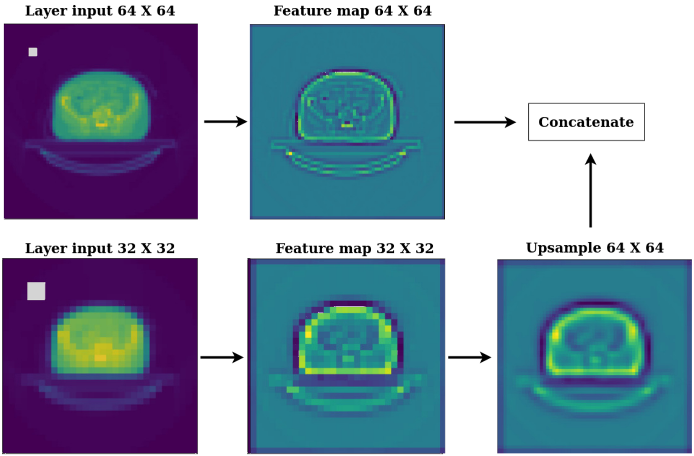


Figure: Multi-resolution analysis: Edge detection shown over encoder-decoder pathway: $\downarrow$ resolution $\implies \uparrow$ relative kernel size (grey) to extract coarser (high level) image features for general localisation. Concatenate with high resolution features for local border segmentation.[6]

[6] Takafumi Nemoto et al. "Efficacy evaluation of 2D, 3D U-Net semantic segmentation and atlas-based segmentation of normal lungs excluding the trachea and main bronchi". In: *Journal of Radiation Research* 61.2 (Feb. 2020), pp. 257–264. ISSN: 1349-9157. DOI: 10.1093/jrr/rrz086

All happy models are alike...

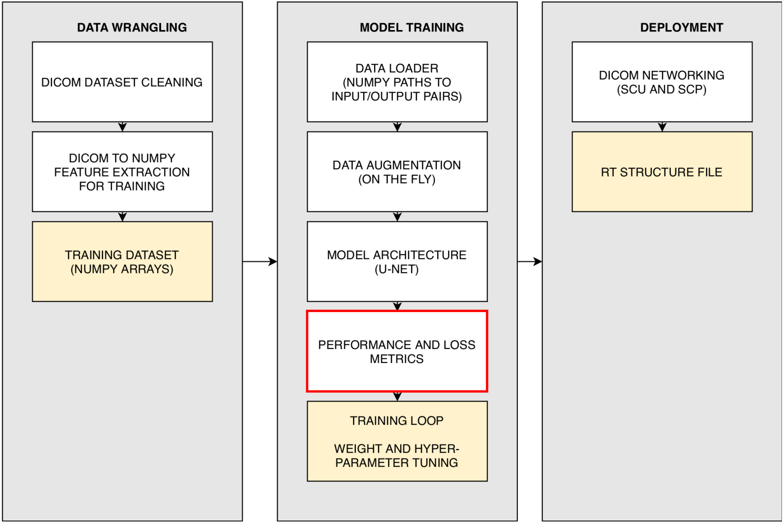


Figure: Modules required for end-to-end deep learning model deployment.

Deployment - DICOM networking

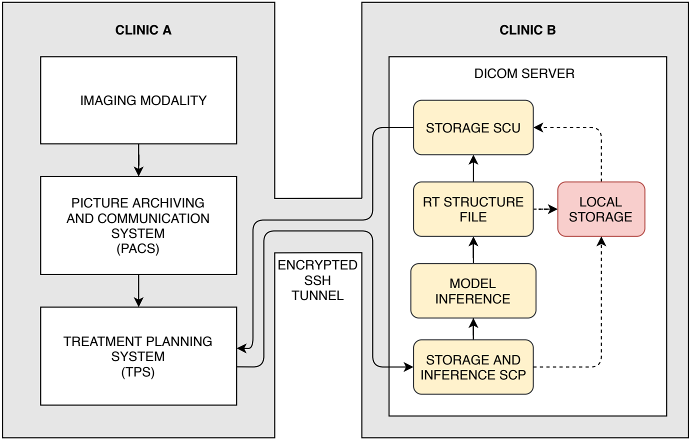


Figure: TPS exports to remote server via DICOM networking protocol.

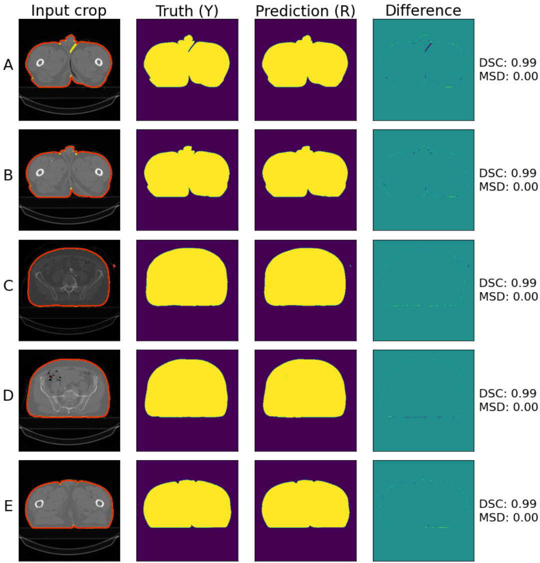


Figure: Representative output for **patient**. Truth contour (yellow), prediction contour (red). Metrics: Dice similarity coefficient (DSC), and mean surface distance (MSD) in mm.

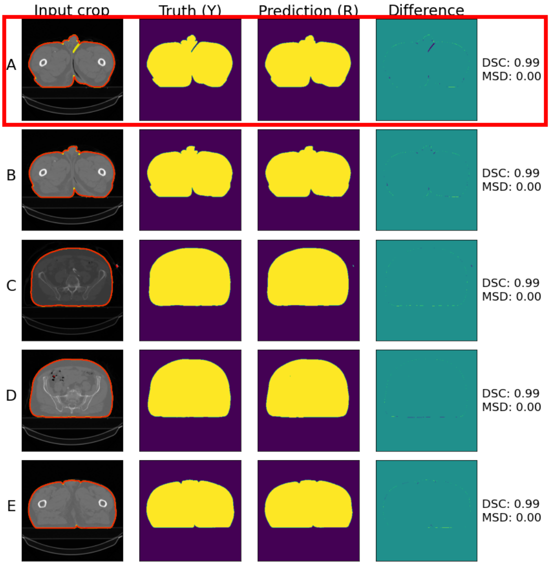


Figure: Representative output for **patient**. Truth contour (yellow), prediction contour (red). Metrics: Dice similarity coefficient (DSC), and mean surface distance (MSD) in mm.

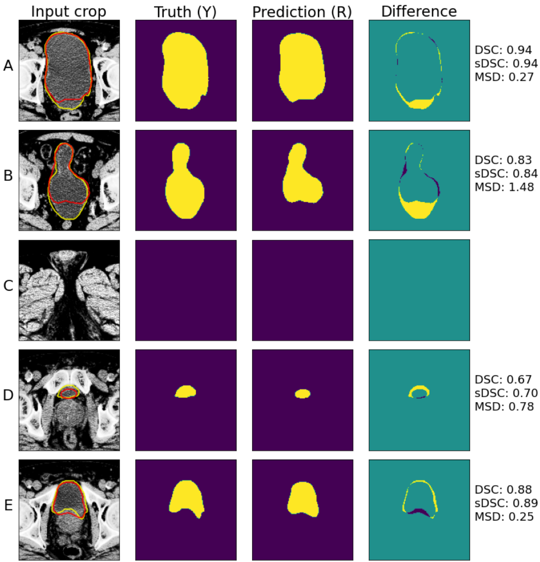


Figure: Representative output for **bladder**. Truth contour (yellow), prediction contour (red). Metrics: Dice coefficient (DSC), surface dice coefficient (sDSC), and mean surface distance (MSD) in mm.

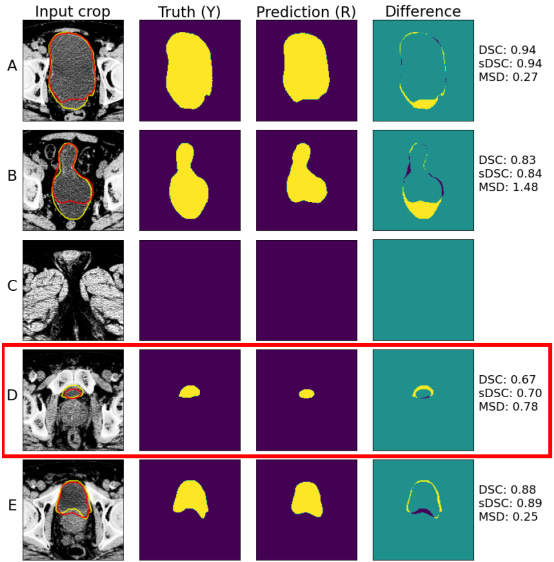


Figure: Representative output for **bladder**. Truth contour (yellow), prediction contour (red). Metrics: Dice coefficient (DSC), surface dice coefficient (sDSC), and mean surface distance (MSD) in mm.

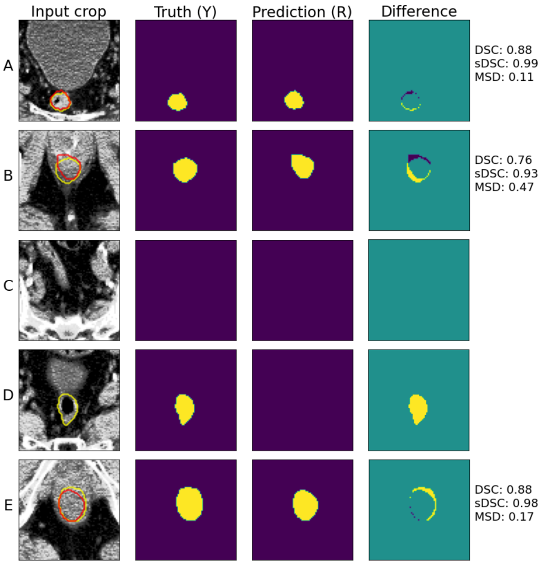


Figure: Representative output for **rectum**. Truth contour (yellow), prediction contour (red). Metrics: Dice coefficient (DSC), surface dice coefficient (sDSC), and mean surface distance (MSD) in mm.

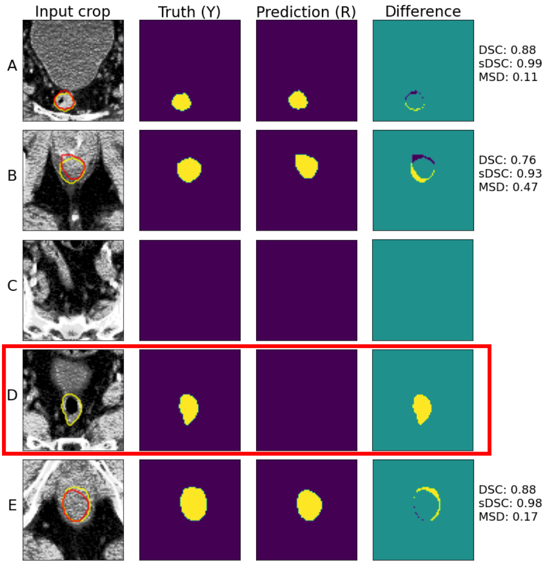


Figure: Representative output for **rectum**. Truth contour (yellow), prediction contour (red). Metrics: Dice coefficient (DSC), surface dice coefficient (sDSC), and mean surface distance (MSD) in mm.

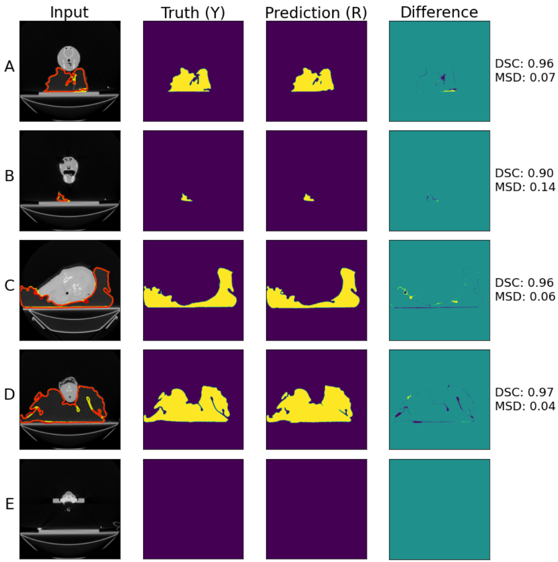


Figure: Representative output for **vacuum bag**. Truth contour (yellow), prediction contour (red). Metrics: Dice similarity coefficient (DSC), and mean surface distance (MSD) in mm.

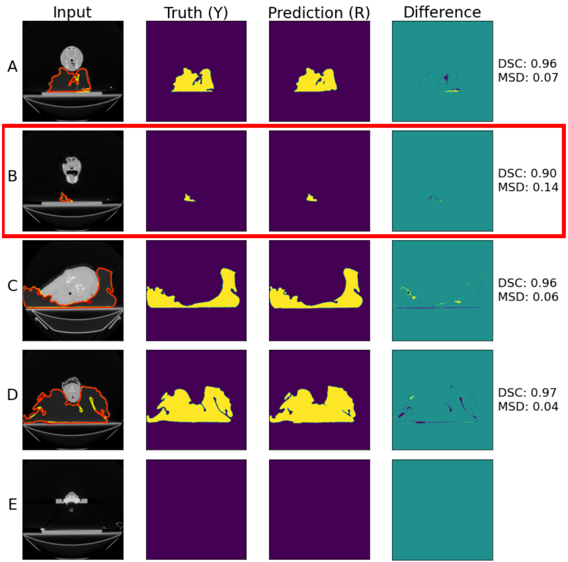


Figure: Representative output for **vacuum bag**. Truth contour (yellow), prediction contour (red). Metrics: Dice similarity coefficient (DSC), and mean surface distance (MSD) in mm.

Structure averaged metrics

Table: Organ specific evaluation on independent dataset.

	sDSC	DSC	MSD (mm)
Pelvic imaging[a]			
Patient		0.99(1)	0.00(5)
Bladder (τ 1.46 mm)	0.9(2)	0.9(2)	1(3)
Rectum (τ 6.99 mm)	0.9(1)	0.7(1)	1(2)
Average		0.9(2)	0.6(2)
Canine imaging			
Vacbag		0.952(1)	0.2(3)

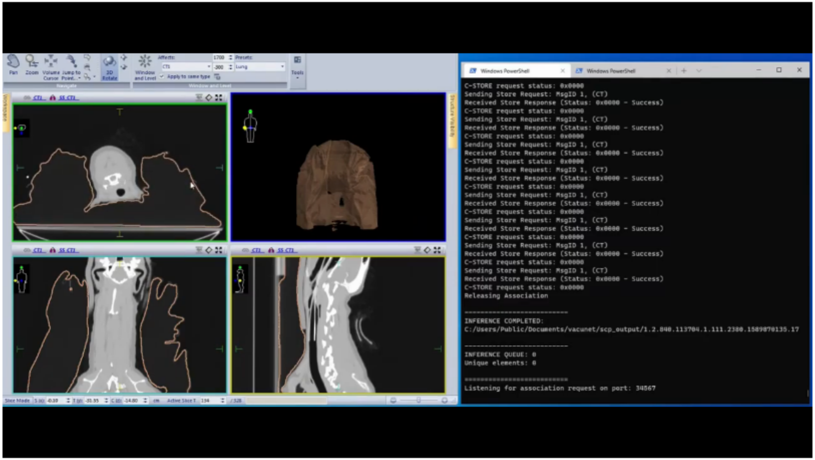
[a] Organ specific tolerance $\tau = \mathrm{MSD}_{95}$ (Top 95% expert performance).
Notation: $\bar{x}(\sigma)$ corresponds to mean $\bar{x}$ with stdev σ in final digit.

Cf. expert IOV.[1]

- Clinically 'acceptable' bladder and rectum DSC $\geq$ 0.7
- Bladder: DSC 0.93(3), MSD 0.9(3) mm.
- Rectum: DSC 0.81(7), MSD 3(2) mm.

[1] Dale Roach et al. "Multi-observer contouring of male pelvic anatomy: Highly variable agreement across conventional and emerging structures of interest". In: *Journal of Medical Imaging and Radiation Oncology* 63.2 (2019), pp. 264–271. DOI: 10.1111/1754-9485.12844

Questions



Video overview: https://youtu.be/fMCv5i6GJWI

References I

Ford, Eric et al. "Strategies for effective physics plan and chart review in radiation therapy: Report of AAPM Task Group 275". In: *Medical Physics* 47.6 (2020), e236–e272. DOI: https://doi.org/10.1002/mp.14030.

Ioffe, Sergey and Christian Szegedy. *Batch Normalization: Accelerating Deep Network Training by Reducing Internal Covariate Shift*. 2015. arXiv: 1502.03167 [cs.LG].

Nemoto, Takafumi et al. "Efficacy evaluation of 2D, 3D U-Net semantic segmentation and atlas-based segmentation of normal lungs excluding the trachea and main bronchi". In: *Journal of Radiation Research* 61.2 (Feb. 2020), pp. 257–264. ISSN: 1349-9157. DOI: 10.1093/jrr/rrz086.

Nikolov, Stanislav et al. *Deep learning to achieve clinically applicable segmentation of head and neck anatomy for radiotherapy*. 2018. arXiv: 1809.04430 [cs.CV].

Roach, Dale et al. "Multi-observer contouring of male pelvic anatomy: Highly variable agreement across conventional and emerging structures of interest". In: *Journal of Medical Imaging and Radiation Oncology* 63.2 (2019), pp. 264–271. DOI: 10.1111/1754-9485.12844.

Ronneberger, Olaf, Philipp Fischer, and Thomas Brox. *U-Net: Convolutional Networks for Biomedical Image Segmentation*. 2015. arXiv: 1505.04597 [cs.CV].

Springenberg, Jost Tobias et al. *Striving for Simplicity: The All Convolutional Net*. 2014. arXiv: 1412.6806 [cs.LG].

Vaassen, Femke et al. "Evaluation of measures for assessing time-saving of automatic organ-at-risk segmentation in radiotherapy". In: *Physics and Imaging in Radiation Oncology* 13 (2020), 1–6. ISSN: 2405-6316. DOI: 10.1016/j.phro.2019.12.001.

References II

Vinod, Shalini K et al. "A review of interventions to reduce inter-observer variability in volume delineation in radiation oncology". In: *Journal of Medical Imaging and Radiation Oncology* 60.3 (2016), pp. 393–406. DOI: 10.1111/1754-9485.12462.