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

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

[1]The University of Sydney (USyd). School of Physics. Institute of Medical Physics.

[2]Riverina Cancer Care Centre (RCCC). Cancer Care Associates.

Thesis: github.com/matthewdeancooper/masters_thesis
Video overview: docs.pymedphys/com/background/autocontouring

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 $\implies$ significant correction times.[3]
- Barrier to future technologies that require fast contouring.[3]

Current deep learning methods

- 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]

Application

Model 1: QA tool - Pelvic imaging (Patient, bladder, rectum).

- Need for delineation to be part of regular QA.[4]
- Alert if prediction differs significantly from expert.

Model 2: Automatic contouring - Canine vacuum bag

- Currently: Manual vacuum bag contouring ($\sim$ 30 min)
- Lower barrier to entry wrt. implementation.

Goal: Performance similar to human experts.

- Performance metric (sDSC) that takes into account expert IOV.[3]
- Stronger correlation with correction time cf. DSC.[5]

---

[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]

[4] 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

[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

Surface dice similarity coefficient (sDSC)

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

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

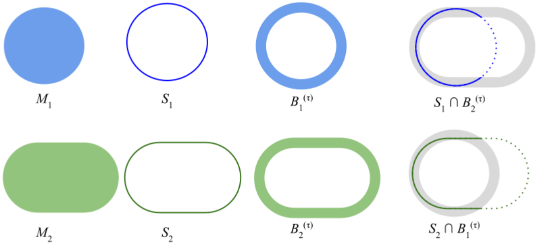


Figure: Clinical performance metric: Illustration of volume masks M_i, surfaces S_i, boundaries $B_i^{(\tau)}$ at organ specific tolerance τ, and intersection of surface boundaries $S_i \cap B_j^{(\tau)}$. Value states the percentage of surface contoured within expert IOV.[3]

[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]

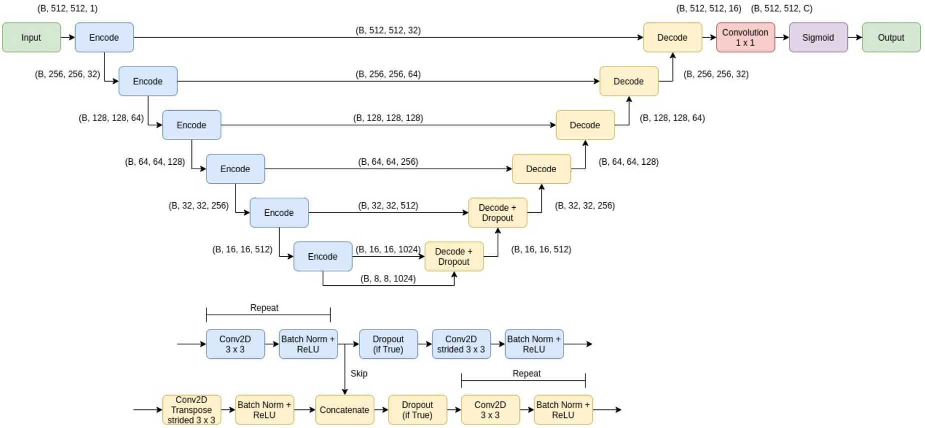


Figure: Modified 2D U-net architecture:[8] Composed of encoding (blue) and decoding blocks (yellow). MaxPooling layers replaced by strided convolution.[11] Added batch normalisation[12] and final sigmoid activation.[3] Tensor dimensions (Batch size, X, Y, Channels).

[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]

All happy models are alike...

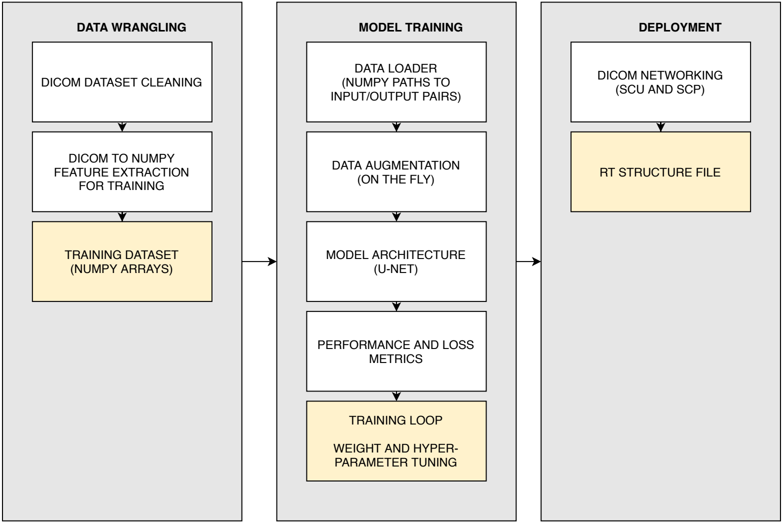


Figure: Modules required for end-to-end deep learning model deployment

Deployment

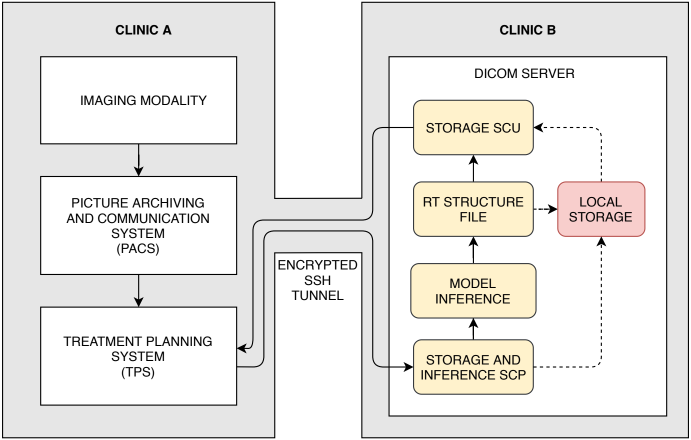


Figure: TPS exports to remote server via DICOM networking protocol

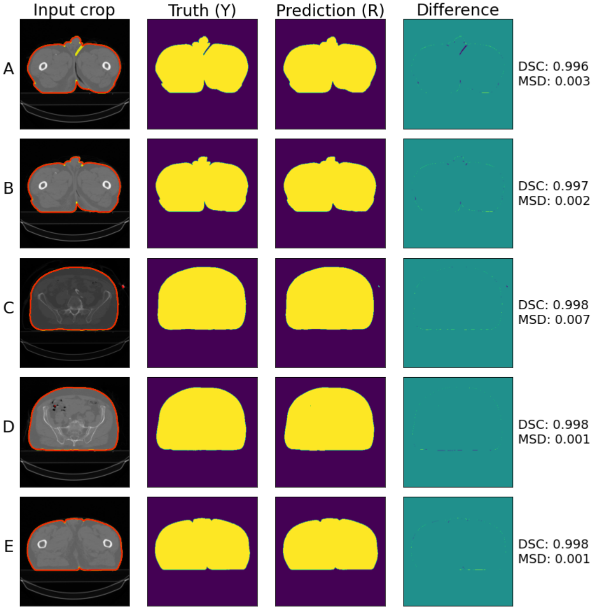


Figure: Representative output for **patient**. Truth contour (yellow), prediction contour (red). Mean surface distance (MSD) mm.

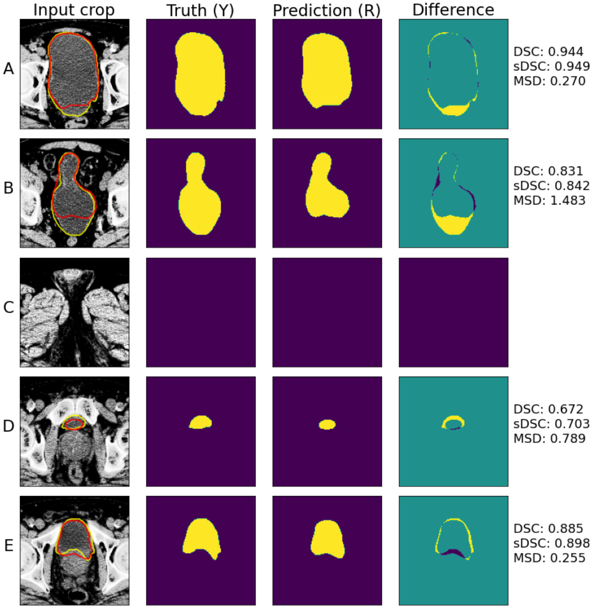


Figure: Representative output for **bladder**. Truth contour (yellow), prediction contour (red). Mean surface distance (MSD) mm. sDSC calculated at τ of 1.46 mm.[2]

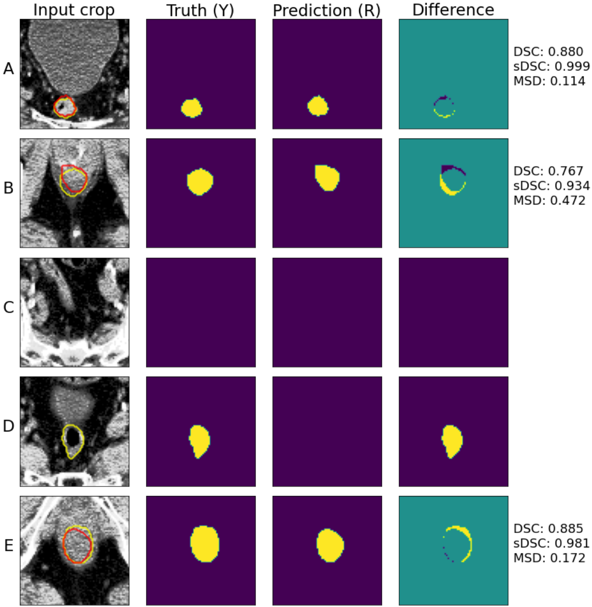


Figure: Representative output for **rectum**. Truth contour (yellow), prediction contour (red). Mean surface distance (MSD) mm. sDSC calculated at τ of 6.99 mm.[2]

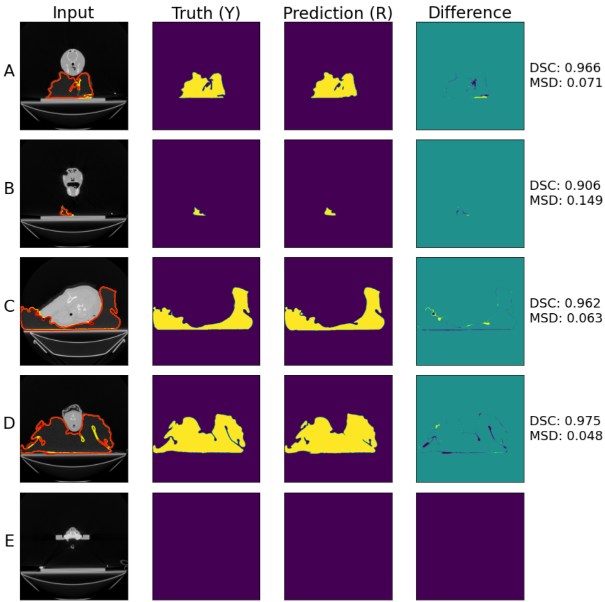


Figure: Representative output for **vacuum bag**. Truth contour (yellow), prediction contour (red). Mean surface distance (MSD) mm.

Structure specific metrics

Table: Organ specific evaluation for proposed models on independent test dataset

	sDSC	DSC	MSD (mm)	Sensitivity	Specificity
Pelvic imaging					
Patient		0.998(0.001)	0.002(0.005)	0.99	0.99
Bladder (τ 1.46 mm)	0.9(0.2)	0.9(0.2)	1(3)	0.79	0.99
Rectum (τ 6.99 mm)	0.9(0.1)	0.7(0.1)	1(2)	0.62	0.99
Average		0.9(0.2)	0.6(2)	0.99	0.99
Canine imaging					
Vacbag		0.952(0.001)	0.2(0.3)	0.95	0.99

Expert IOV.[2]

- Clinically 'acceptable' bladder and rectum DSC ≥ 0.7
- Bladder: DSC 0.93 ± 0.03, MSD 0.9(0.3) mm.
- Rectum: DSC 0.81 ± 0.07, MSD 3(2) mm.
- Organ specific tolerance $\tau = \text{MSD}_{95}$ (ie. Top 95% expert performance).[3]

[2] 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

[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]

Conclusion

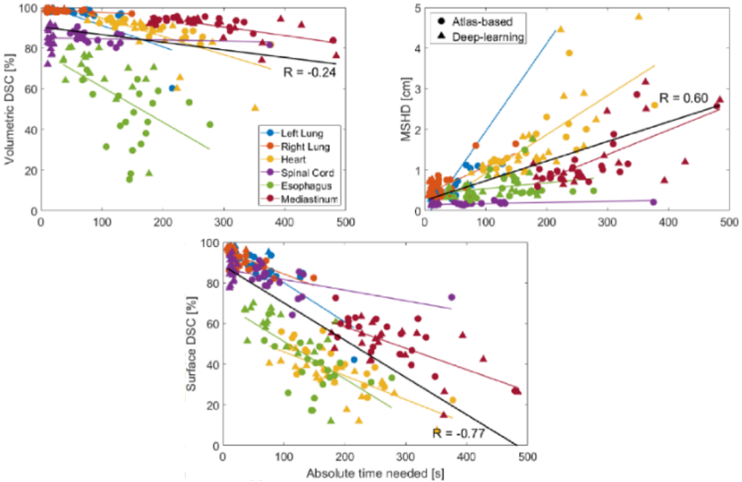
Pelvic imaging model:

- Patient contouring within tolerances (DSC 0.998).
- Suspect more data will improve bladder and rectum volumes (DSC 0.860, 0.670).
  - cf. S.O.T.A commercial DL solution (0.97, 0.79).[15]
- Weighted soft DSC loss significantly improved performance on class imbalanced data.

Canine imaging model:

- Successfully deployed to clinic under a prototype warning
- Performance improvement of approximately 30 minutes per patient

[15] Jordan Wong et al. "Comparing deep learning-based auto-segmentation of organs at risk and clinical target volumes to expert inter-observer variability in radiotherapy planning". In: *Radiotherapy and Oncology* 144 (2020), 152–158. DOI: 10.1016/j.radonc.2019.10.019



Comparison of common segmentation metrics with surface DSC (sDSC) for ability to infer absolute time required for automatic contour correction.[18]

[18] 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

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].

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.

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.

Wong, Jordan et al. "Comparing deep learning-based auto-segmentation of organs at risk and clinical target volumes to expert inter-observer variability in radiotherapy planning". In: *Radiotherapy and Oncology* 144 (2020), 152–158. DOI: 10.1016/j.radonc.2019.10.019.