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

Video overview: docs.pymedphys/com/background/autocontouring

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]
- Barrier to future technologies that require fast contouring.[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 - Pelvic imaging for prostate cancer.

- Contour patient, bladder, rectum volumes
- Alert if prediction differs significantly from expert.
- Need for delineation to be part of regular QA.[4]

Model 2: Automatic contouring - 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.[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]

[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

Performance - Surface dice similarity coefficient (sDSC)

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

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

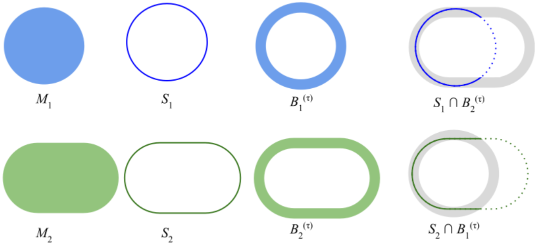

Figure: Clinical performance metric: Illustration of volume masks M_i, surfaces S_i, boundaries $B_i^{(\tau)}$ at organ specific tolerance τ, and intersection of surface boundaries $S_i \cap B_j^{(\tau)}$. sDSC is the percentage of surface contoured within expert IOV.[3]

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

Modules - All happy models are alike...

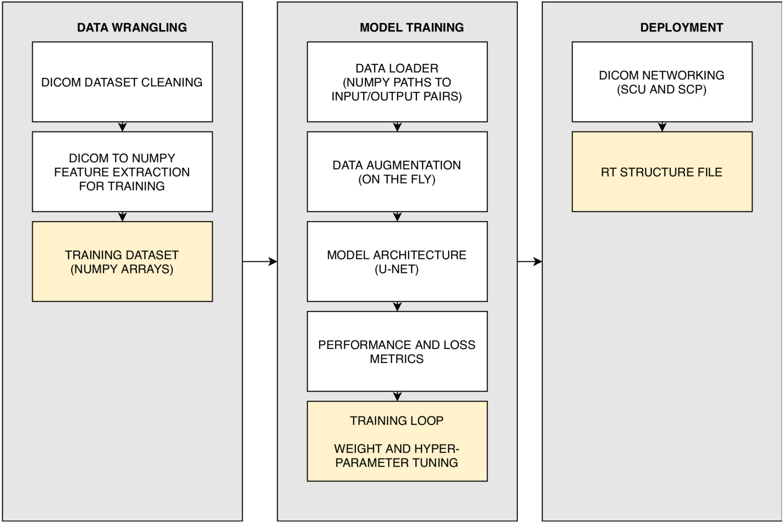

Figure: Modules required for end-to-end deep learning model deployment

Modules - All happy models are alike...

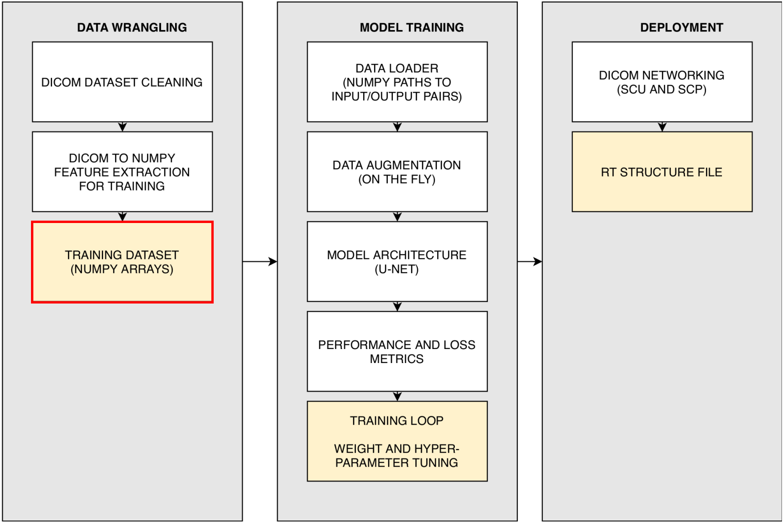

Figure: Modules required for end-to-end deep learning model deployment

Modules - All happy models are alike...

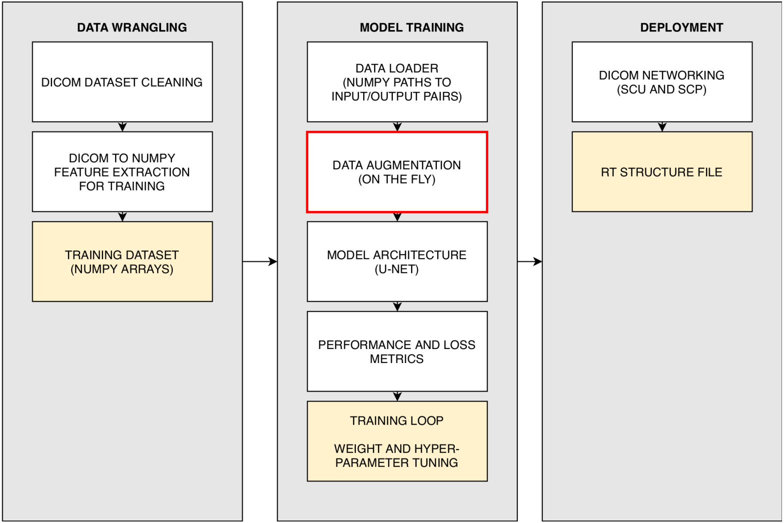

Figure: Modules required for end-to-end deep learning model deployment

Modules - All happy models are alike...

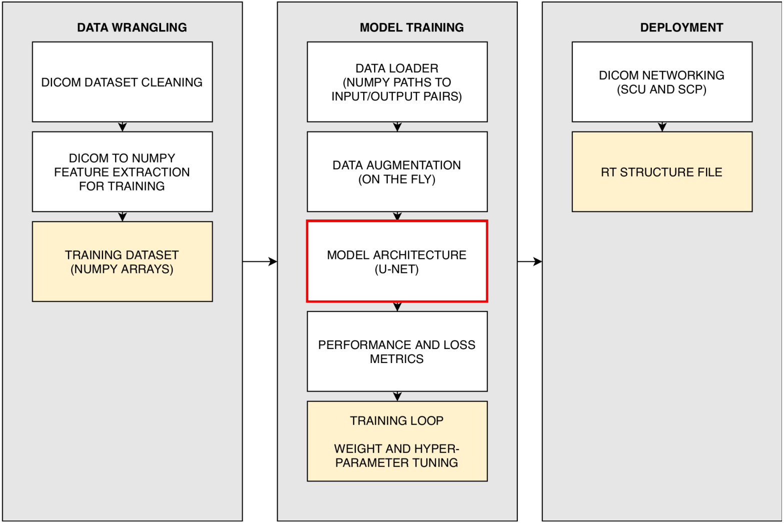

Figure: Modules required for end-to-end deep learning model deployment

Modules - All happy models are alike...

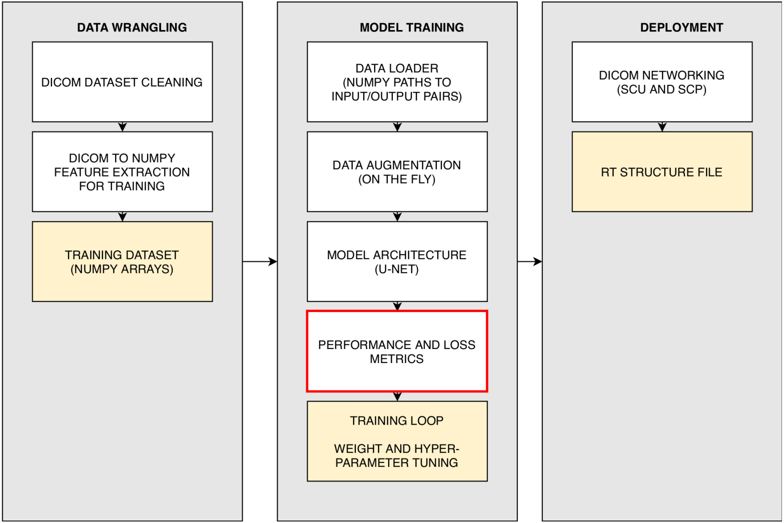

Figure: Modules required for end-to-end deep learning model deployment

Deployment - DICOM networking

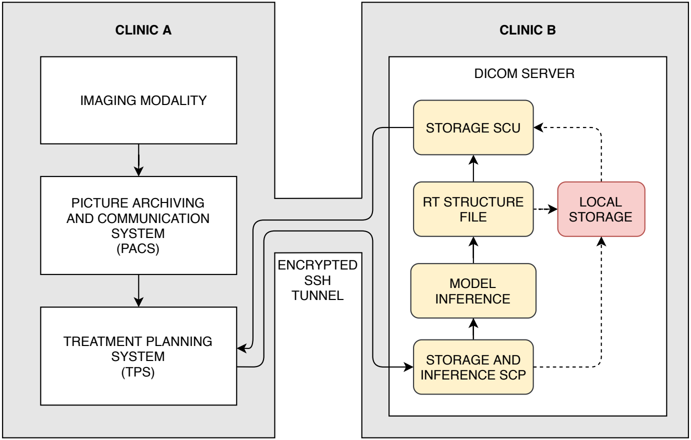

Figure: TPS exports to remote server via DICOM networking protocol

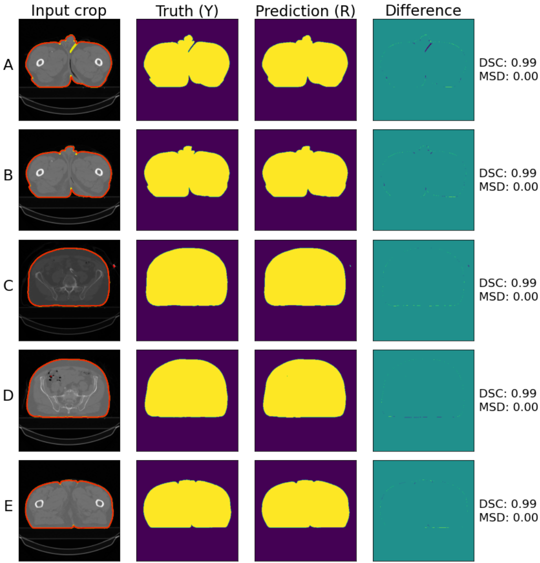

Figure: Representative output for **patient**. Truth contour (yellow), prediction contour (red). Metrics: Dice similarity coefficient (DSC), and mean surface distance (MSD) in mm.

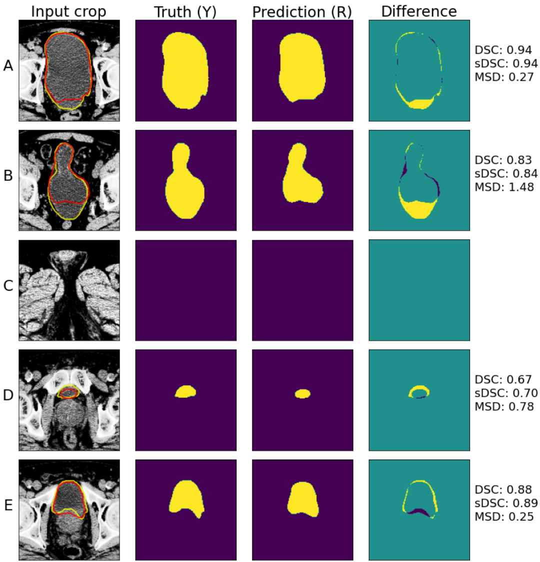

Figure: Representative output for **bladder**. Truth contour (yellow), prediction contour (red). Metrics: Dice coefficient (DSC), surface dice coefficient (sDSC), and mean surface distance (MSD) in mm

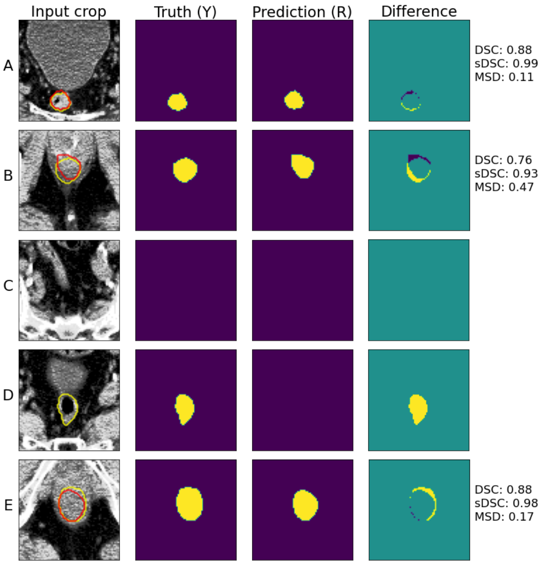

Figure: Representative output for **rectum**. Truth contour (yellow), prediction contour (red). Metrics: Dice coefficient (DSC), surface dice coefficient (sDSC), and mean surface distance (MSD) in mm

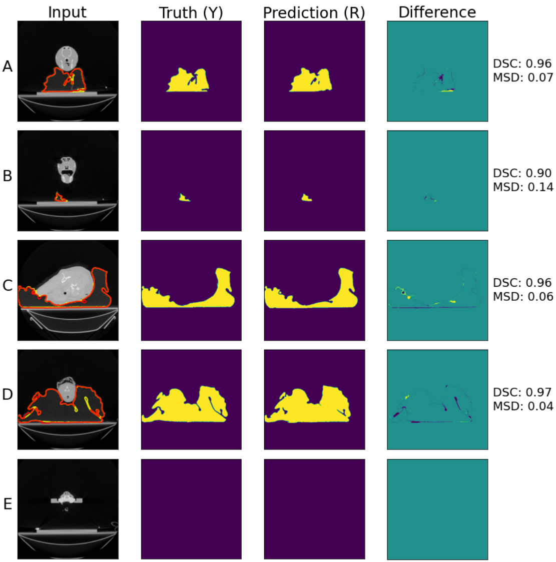

Figure: Representative output for **vacuum bag**. Truth contour (yellow), prediction contour (red). Metrics: Dice similarity coefficient (DSC), and mean surface distance (MSD) in mm.

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)

Cf. expert IOV[2]

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

[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 and future work

Pelvic imaging model:

- Patient contouring within tolerances.
- Suspect more data will improve bladder and rectum segmentation.
- 3D architecture may identify gaseous rectal volumes.

Canine imaging model:

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

Future

- Develop a continuous valued surrogate for sDSC to optimise directly.

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.

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.

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.