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

Docs: docs.pymedphys/com/background/autocontouring

Introduction: Motivation

Variability

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

Time constraints

- Manual $\implies$ $\leq$ 4 hrs (head & neck).[3]
- Atlas methods $\implies$ significant correction times.[4]
- Barrier to future technologies that require fast contouring.[4]

Current deep learning methods

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

[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. eprint: https://aapm.onlinelibrary.wiley.com/doi/pdf/10.1002/mp.14030. URL: https://aapm.onlinelibrary.wiley.com/doi/abs/10.1002/mp.14030

[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

Introduction: Clinical application

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

- Need for delineation to be part of regular QA.[4]
- Potential to manage some prominent hazards identified by the task group TG275
- Alert if prediction differs significantly from expert.

Model 2: Automatic contouring (SASH) - Canine vacuum bag

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

Goal: Achieve 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. URL: http://dx.doi.org/10.1016/j.phro.2019.12.001

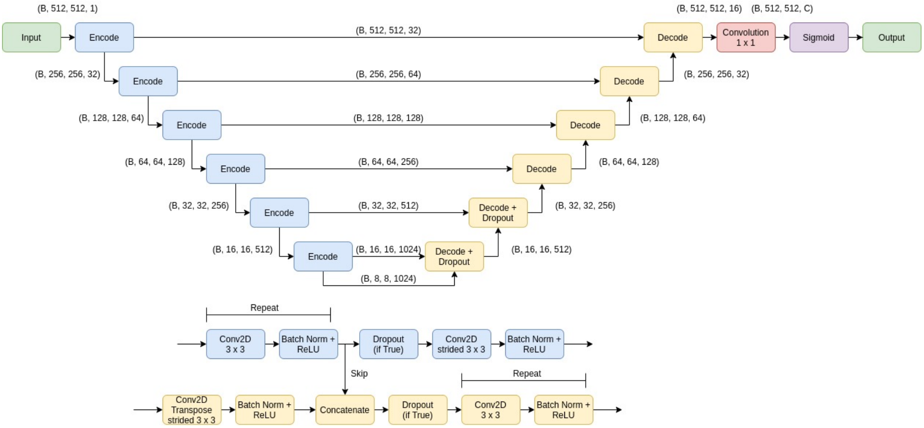


Figure: Modified 2D U-net architecture:[8] Composed of encoding (blue) and decoding blocks (yellow). MaxPooling layers replaced by strided convolution[11]. Added batch normalisation[12] and final sigmoid activation.[3] Tensor dimensions (Batch size, X, Y, Channels).

[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?

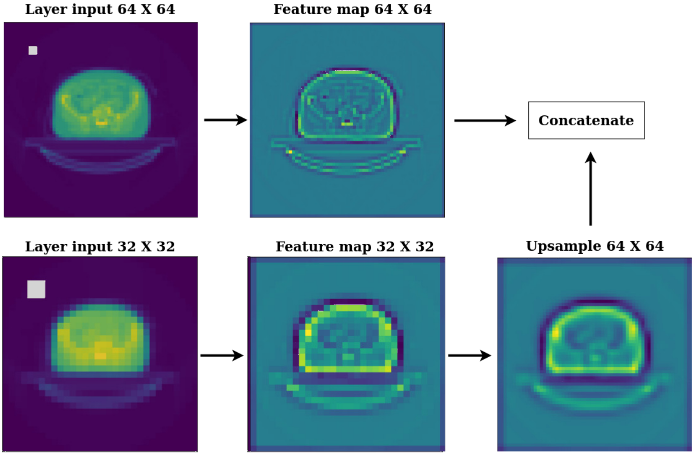


Figure: Multi-resolution analysis: Edge detection shown over encoder-decoder pathway: $\downarrow$ resolution $\implies \uparrow$ relative kernel size (grey) to extract coarser (high level) image features for general localisation. Concatenate with high resolution features for local border segmentation.[6]

[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

Code: All happy models are alike...

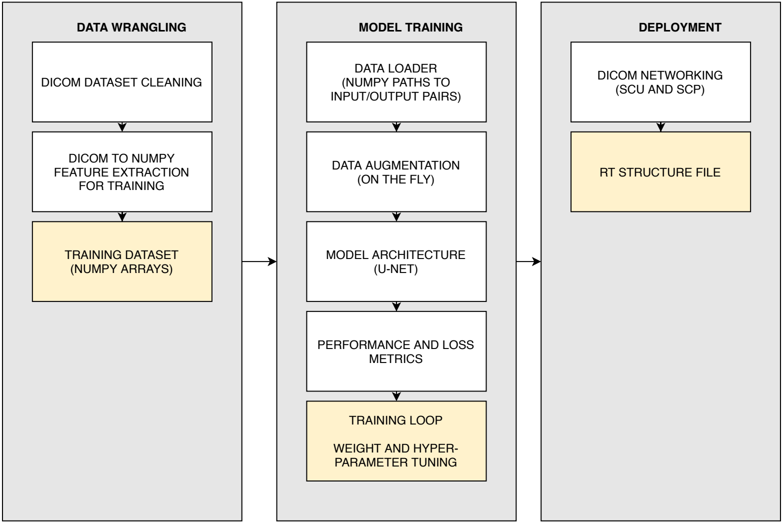


Figure: Modules required for end-to-end deep learning model deployment

Code: Deployment loop

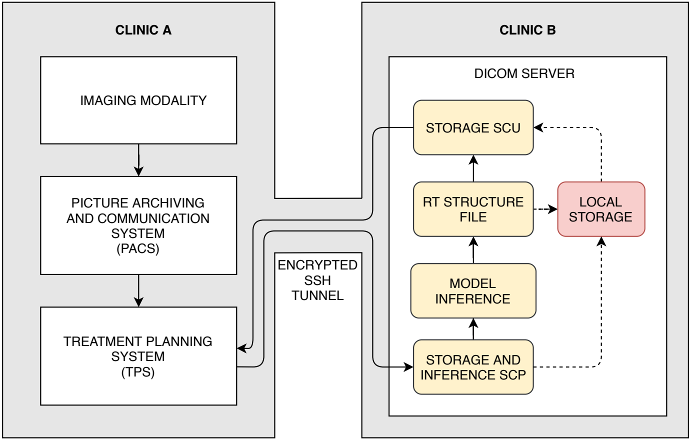
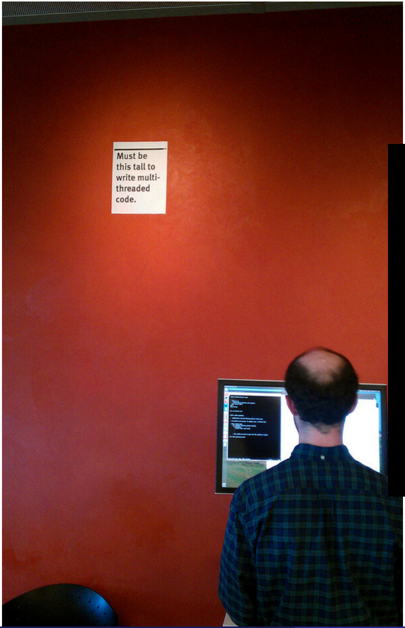
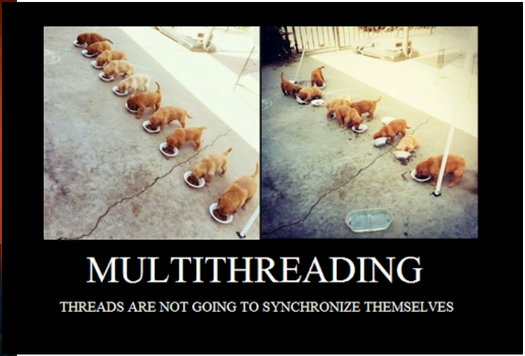


Figure: Implementation of DICOM network - demonstration!

Code: Forced multi-threading





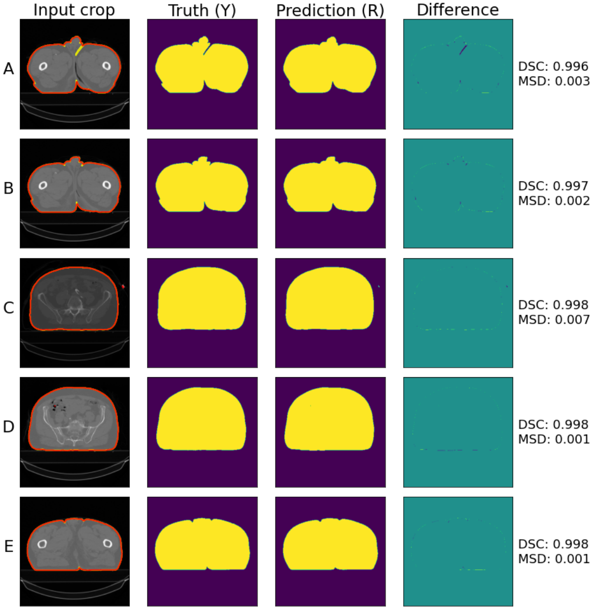


Figure: Representative output for **patient**. Truth contour (yellow), prediction contour (red). Mean surface distance (MSD) mm.

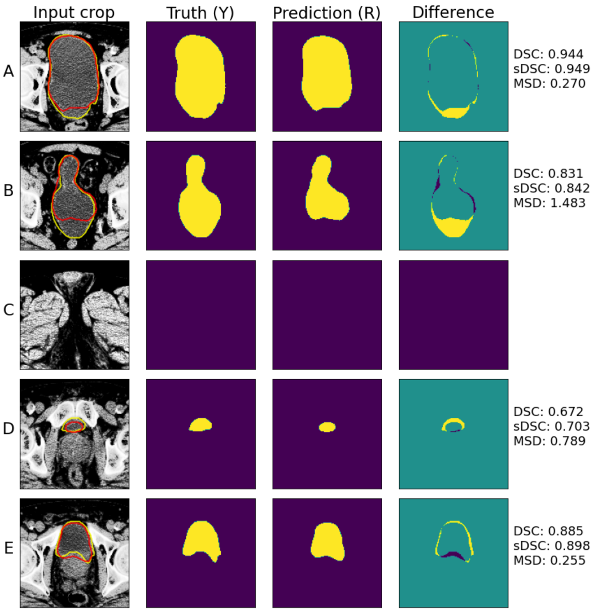


Figure: Representative output for **bladder**. Truth contour (yellow), prediction contour (red). Mean surface distance (MSD) mm. sDSC calculated at τ of 1.46 mm.[2]

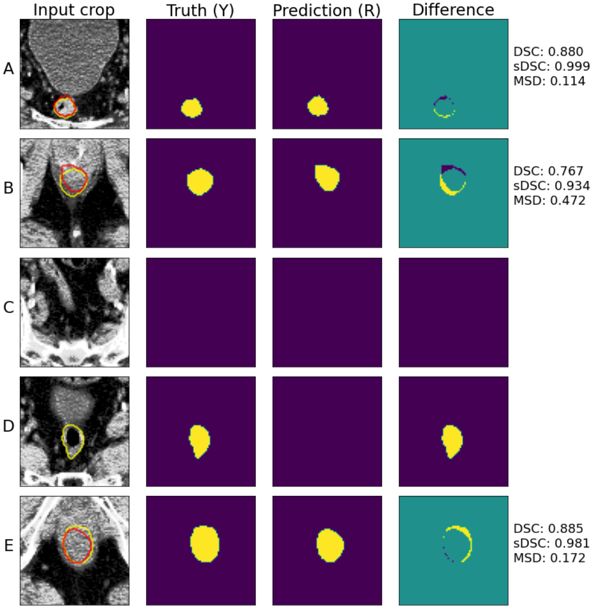


Figure: Representative output for **rectum**. Truth contour (yellow), prediction contour (red). Mean surface distance (MSD) mm. sDSC calculated at τ of 6.99 mm.[2]

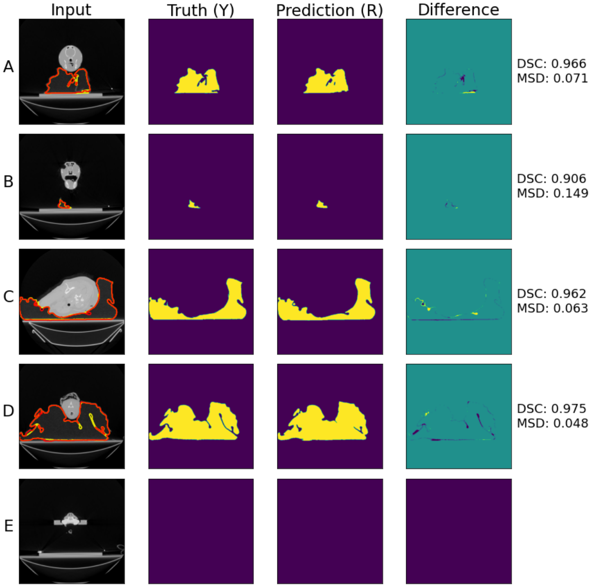


Figure: Representative output for **vacuum bag**. Truth contour (yellow), prediction contour (red). Mean surface distance (MSD) mm.

Discussion: Structure specific metrics

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

Organ: Mean(Std)	sDSC (τ)	DSC	MSD (mm)	Sensitivity	Specificity
Pelvic imaging					
Patient		0.998(0.001)	0.002(0.005)	0.997	0.999
Bladder (τ 1.46 mm)	0.9(0.2)	0.9(0.2)	1(3)	0.786	0.999
Rectum (τ 6.99 mm)	0.9(0.1)	0.7(0.1)	1(2)	0.619	0.999
Average		0.9(0.2)	0.6(2)	0.991	0.999
Canine imaging					
Vacbag		0.952(0.001)	0.2(0.3)	0.953	0.995

- Organ specific tolerance $\tau = \mathrm{MSD}_{95}$ (ie. Top 95% expert performance).[3]

Cf. Experts IOV.[2]

- Clinically 'acceptable' bladder and rectum DSC ≥ 0.7
- Bladder: DSC 0.93 ± 0.03, MSD 0.99(0.30) mm.
- Rectum: DSC 0.81 ± 0.07, MSD 2.862(2.066) 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: Summary

Pelvic imaging model:

- Patient contouring within tolerances (DSC 0.998).
- Suspect more data will improve bladder and rectum volumes (DSC 0.860, 0.670).
  S.O.T.A commercial DL solution (0.97, 0.79).[15]
  S.O.T.A open-source DL solution (0.95, 0.92).[16]
- 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. ISSN: 0167-8140. DOI: 10.1016/j.radonc.2019.10.019. URL: http://dx.doi.org/10.1016/j.radonc.2019.10.019

[16] Samaneh Kazemifar et al. "Segmentation of the prostate and organs at risk in male pelvic CT images using deep learning". In: *Biomedical Physics and Engineering Express* 4.5 (2018), p. 055003. ISSN: 2057-1976. DOI: 10.1088/2057-1976/aad100. URL: http://dx.doi.org/10.1088/2057-1976/aad100

Conclusion: Future work

Post-project goals:

- Need for more data - limited dataset reduces generalisability.[17]
  - S.O.T.A open-source: 600 - 1000 patients cf. 15.[3]
- sDSC is a 'hard' metric. Developing a soft surrogate may allow for the direct optimisation of this metric
- Potential improvements under more complicated architecture

---

[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]

[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

[17] D. Shen, G. Wu, and H. I. Suk. "Deep Learning in Medical Image Analysis". In: *Annu Rev Biomed Eng* 19 (June 2017). [DOI:10.1146/annurev-bioeng-071516-044442], pp. 221–248

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. eprint: https://aapm.onlinelibrary.wiley.com/doi/pdf/10.1002/mp.14030. URL: https://aapm.onlinelibrary.wiley.com/doi/abs/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].

Kazemifar, Samaneh et al. "Segmentation of the prostate and organs at risk in male pelvic CT images using deep learning". In: *Biomedical Physics and Engineering Express* 4.5 (2018), p. 055003. ISSN: 2057-1976. DOI: 10.1088/2057-1976/aad100. URL: http://dx.doi.org/10.1088/2057-1976/aad100.

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].

References II

Shen, D., G. Wu, and H. I. Suk. "Deep Learning in Medical Image Analysis". In: *Annu Rev Biomed Eng* 19 (June 2017). [DOI:10.1146/annurev-bioeng-071516-044442], pp. 221–248.

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. URL: http://dx.doi.org/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. ISSN: 0167-8140. DOI: 10.1016/j.radonc.2019.10.019. URL: http://dx.doi.org/10.1016/j.radonc.2019.10.019.