Uncertainty in multitask learning: joint representations for probabilistic MR-only radiotherapy treatment planning

Felix J.S. Bragman[1], Ryutaro Tanno[1], Zach Eaton-Rosen[1], Wenqi Li[1]
David J. Hawkes[1], Sébastien Ourselin[2], Daniel C. Alexander[1,3], Jamie R. McClelland[1] and M. Jorge Cardoso[2,1]

1. Centre for Medical Image Computing, University College London
2. Biomedical Engineering and Imaging Sciences, King's College London
3. Clinical Imaging Research Centre, National University of Singapore

Motivation and Overview

- **Multi-task learning** is dependent on the relative weighting of task losses and the mechanism for sharing network weights.
- Task loss weightings are generally hyper-parameters or learned [1]
- We propose a **probabilistic multi-task network** (Fig. 1) that:
    - a) estimates **heteroscedastic uncertainty** for spatially adaptive task loss weighting
    - b) captures **model uncertainty** through approximate Bayesian inference
- We apply our model to **MR-only radiotherapy treatment planning**,
- **Results** show:
    1. heteroscedastic uncertainty **improves** multi-task learning over learned task loss weights
    2. the estimated uncertainty can be exploited for **quality assurance and control** of the network

MR-only radiotherapy treatment planning

- Joint synthesis of a CT scan (synCT) - **regression** - and localisation of organs at risk (OAR) **- segmentation -** from an input MRI scan

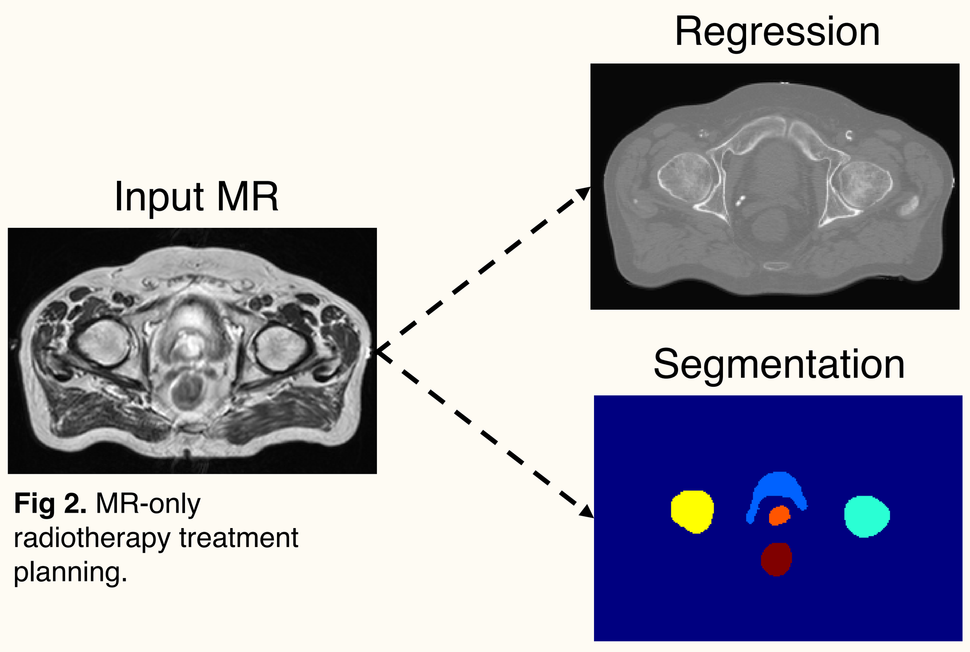

Fig 2. MR-only radiotherapy treatment planning.

- Previous synthesis methods [2, 3] do not jointly segment OARs
- CT synthesis methods are generally fully deterministic
- A system that can sample both synCT and OAR segmentations will enable end-to-end uncertainty aware probabilistic planning

Probabilistic dual-task neural network

- Probabilistic multi-task learning with hard-parameter sharing: representation network + task-specific networks

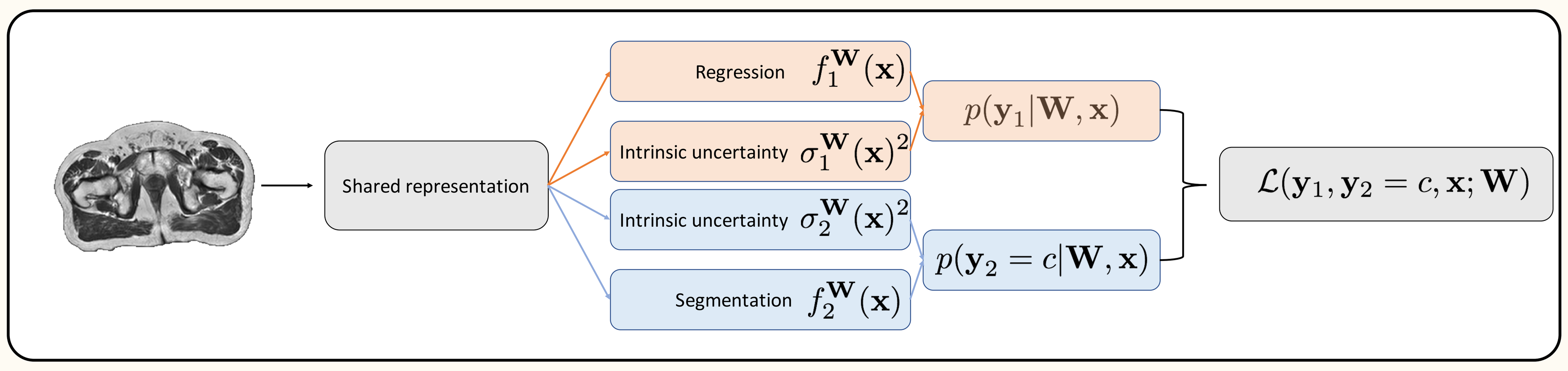

Fig 1. A representation network learns an invariant feature space for the anatomy. A task-specific network then learns the non-linear mapping between the input and the various output tasks. The task-specific likelihoods $p(\mathbf{y}_i|\mathbf{W},\mathbf{x})$ are combined to yield the multi-task likelihood.

Task weighting with heteroscedastic uncertainty

- Heteroscedastic uncertainty represents inherent ambiguity present in the MR-CT intensity mapping and in obtaining voxel-wise class memberships
- Uncertainty is spatially varying and task dependent – this is exploited as a mechanism for task loss weighting

❑ **Regression noise model**

Likelihood function

$$p\left(\mathbf{y}_1|\mathbf{W},\mathbf{x}\right)=\mathcal{N}(f_1^{\mathbf{W}}(\mathbf{x}),\sigma_1^{\mathbf{W}}(\mathbf{x})^2)$$

❑ **Segmentation noise model**

Likelihood function

$$p\left(\mathbf{y}_2|\mathbf{W},\mathbf{x}\right)=\mathrm{Softmax}\big(f_2^{\mathbf{W}}(\mathbf{x})/2\sigma_2^{\mathbf{W}}(\mathbf{x})^2\big)$$

❑ **Multi-task heteroscedastic likelihood**

$$\mathcal{L}(\mathbf{y}_1,\mathbf{y}_2=c,\mathbf{x};\mathbf{W})=\frac{||\mathbf{y}_1-f_1^{\mathbf{W}}(\mathbf{x})||^2}{2\sigma_1^{\mathbf{W}}(\mathbf{x})^2}+\frac{\mathrm{CE}(f_2^{\mathbf{W}}(\mathbf{x}),\mathbf{y}_2=c)}{2\sigma_2^{\mathbf{W}}(\mathbf{x})^2}+\log\left(\sigma_1^{\mathbf{W}}(\mathbf{x})^2\sigma_2^{\mathbf{W}}(\mathbf{x})^2\right)$$

Model uncertainty

- We account for parameter uncertainty through an approximation of the posterior distribution over the weights
- The posterior distribution is approximated through Bernouilli binary dropout [4]

Posterior distribution approximation

$$q(\mathbf{W})\approx p(\mathbf{W}|\mathbf{X},\mathbf{Y_1},\mathbf{Y_2})$$

Sampling posterior during training

$$w'\sim q(\mathbf{W})\quad \mathbf{f}^{w'}(\mathbf{x}):=[f_1^{w'}(\mathbf{x}),f_2^{w'}(\mathbf{x}),\sigma_1^{w'}(\mathbf{x})^2,\sigma_2^{w'}(\mathbf{x})^2]$$

Quantifying total uncertainty over predictions

- At test time, for each input patch $\mathbf{x}$, output samples $\{\mathbf{f}^{w^{(t)}}(\mathbf{x})\}_{t=1}^{\mathrm{T}}$ are obtained by performing T stochastic passes through the network such that $\{w^{(t)}\}_{t=1}^{\mathrm{T}}\sim q(\mathbf{W})$
- The variance of the predictive distribution quantifies the **predictive uncertainty**
- We use the **predictive mean** as final estimates for $\mathbf{f}^{w}(\mathbf{x})$ whilst the **total uncertainty** is the sum of the predictive uncertainty and modelled heteroscedastic noise.

Model performance

- We tested on 15 prostate scans using 3-fold cross-validation
- We trained the network on randomly sampled axial slices

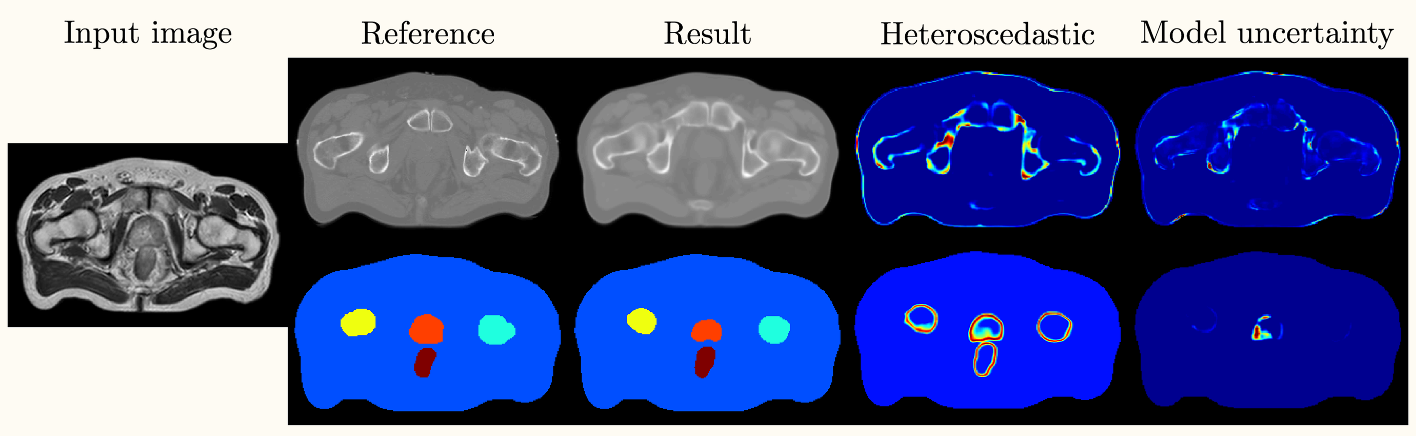

Fig 3. Example of the model output. For each new subject, we obtain: 1) a synCT, 2) segmentation of the organs, 3) the uncertainty across both predictions.

- Joint modelling of heteroscedastic and parameter uncertainty **achieves best performance** on synCT regression and **outperforms** homoscedastic task weighting
- Equivalent results in segmentation with the state of the art in pelvic segmentation [5]

Models	All	Bone	L femur	R femur	Prostate	Rectum	Bladder
Regression - synCT - Mean Absolute Error (HU)							
HighResNet [7]	**48.1(4.2)**	**131(14.0)**	78.6(19.2)	**80.1(19.6)**	**37.1(10.4)**	63.3(47.3)	**24.3(5.2)**
HighResNet + dropout	**47.4(3.0)**	**130(12.1)**	78.0(14.8)	77.0(13.0)	**36.5(7.8)**	67(44.6)	**24.1(7.5)**
HighResNet + dropout + hetero [6]	44.5(3.6)	128(17.1)	75.8(20.1)	74.2(17.4)	31.2(7.0)	56.1(45.5)	17.8(4.7)
Multi-task + homo noise weighting [1]	44.3(3.1)	126(14.4)	74.0(19.5)	73.7(17.1)	29.4(4.7)	58.4(48.0)	18.2(3.5)
Multi-atlas propagation [5]	45.7(4.6)	125(10.3)	-	-	-	-	-
Multi-task + dropout + hetero	43.3(2.9)	121(12.6)	69.7(13.7)	67.8(13.2)	28.9(2.9)	55.1(48.1)	18.3(6.1)
Segmentation - OAR - Fuzzy DICE score							
HighResNet [7]	-	-	0.91(0.02)	0.90(0.04)	0.67(0.12)	0.70(0.15)	0.92(0.05)
HighResNet + dropout	-	-	0.85(0.03)	0.90(0.04)	0.66(0.12)	0.69(0.13)	0.90(0.07)
HighResNet + dropout + hetero [6]	-	-	0.92(0.02)	0.92(0.01)	0.77(0.07)	0.74(0.13)	0.92(0.03)
Multi-task + homo noise weighting [1]	-	-	0.92(0.02)	0.92(0.02)	0.73(0.07)	0.76(0.10)	0.93(0.02)
Multi-atlas propagation [5]	-	-	0.89(0.02)	0.90(0.01)	0.73(0.06)	0.77(0.06)	0.90(0.03)
Multi-task + dropout + hetero	-	-	0.91(0.02)	0.91(0.02)	0.70(0.06)	0.74(0.12)	0.93(0.04)

Uncertainty as a quality control mechanism

- Total uncertainty correlates with regression errors in the synCT

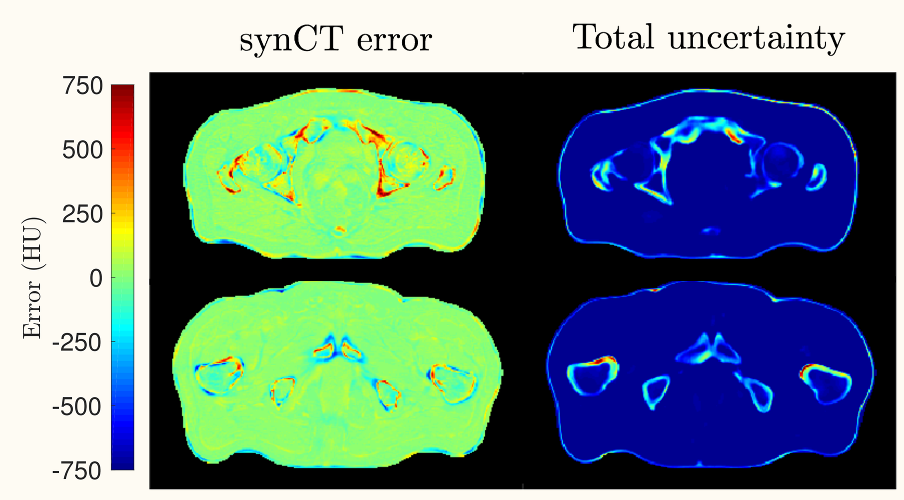

Fig 4. Correlation between total uncertainty and regression error.

- Our method produces well calibrated uncertainty measures

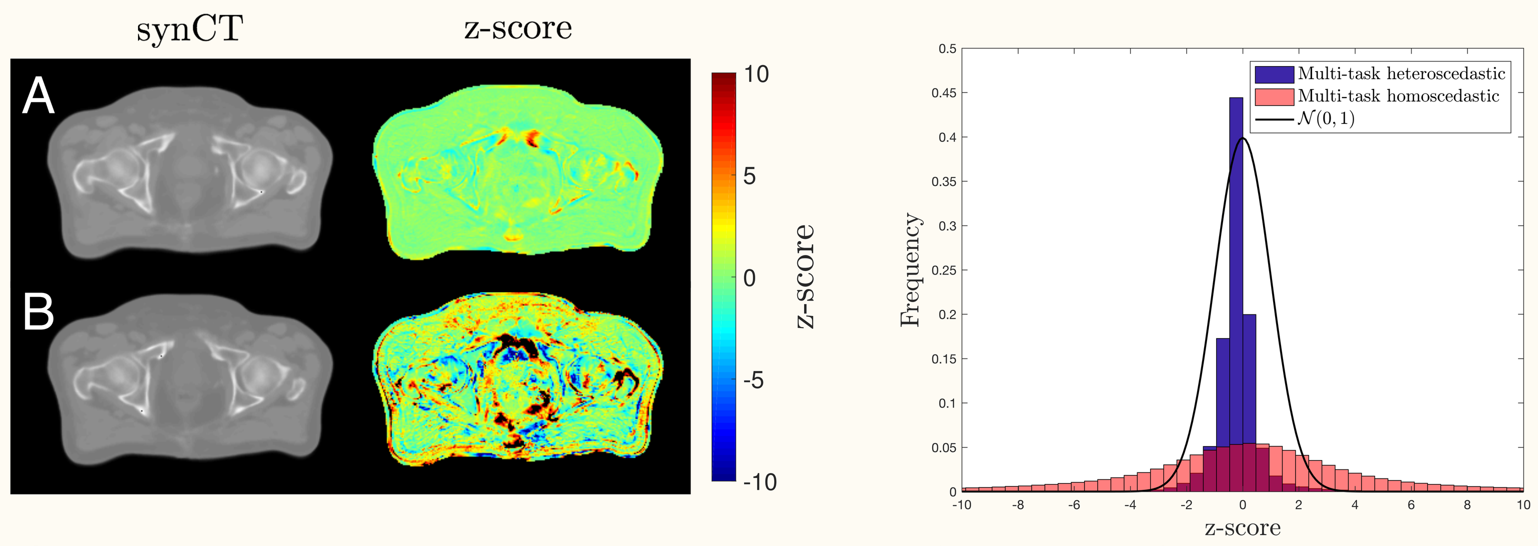

Fig 5. We calculated the z-score using the total uncertainty estimated in a) our model and b) with homoscedastic task weighting.

References

1. A. Kendall et al. In *CVPR*, 2018
2. D. Nie et al. arXiv:1612.05362
3. J. Wolterink et al., In *SASHIMI,* 2017
4. Y. Gal and Z. Ghahramani. In *ICML*, 2016
5. N. Burgos et al. In *Phys. Med. Bio.* 2017
6. A. Kendall et al. In *NIPS*, 2017
7. W. Li et al. In *IPMI*, 348-360, 2017

Acknowledgments. This work was primarily supported by the CRUK Accelerator Grant A21993. RT was supported by Microsoft Scholarship. ZT was supported by an EPSRC Doctoral Prize. DA was supported by EU Horizon 2020 Research and Innovation Programme Grant 666992 and EPSRC Grant M020533, M006093 and J020990. We thank NVIDIA corporation for the hardware donation.

Centre for Medical Image Computing

CANCER RESEARCH UK