

# Quality control in radiotherapy treatment planning using multi-task learning and uncertainty estimation

Felix J.S. Bragman<sup>1</sup>, Ryutaro Tanno<sup>1</sup>, Zach Eaton-Rosen<sup>1</sup>, Wenqi Li<sup>1</sup>

David J. Hawkes<sup>1</sup>, Sébastien Ourselin<sup>2</sup>, Daniel C. Alexander<sup>1,3</sup>, Jamie R. McClelland<sup>1</sup> and M. Jorge Cardoso<sup>2,1</sup>

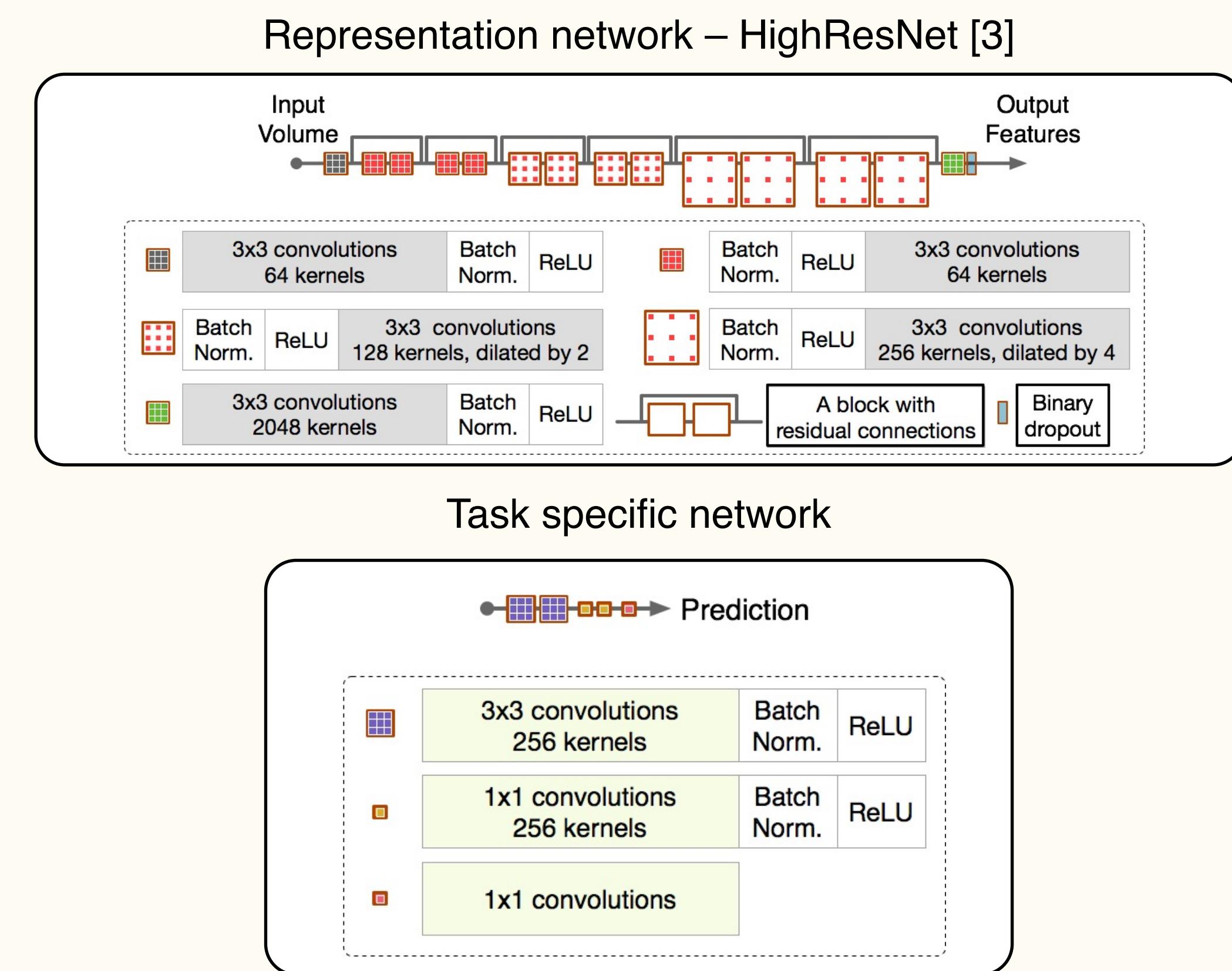
1. Centre for Medical Image Computing, University College London
2. Biomedical Engineering and Imaging Sciences, King's College London
3. Clinical Imaging Research Centre, National University of Singapore



## Motivation and Overview

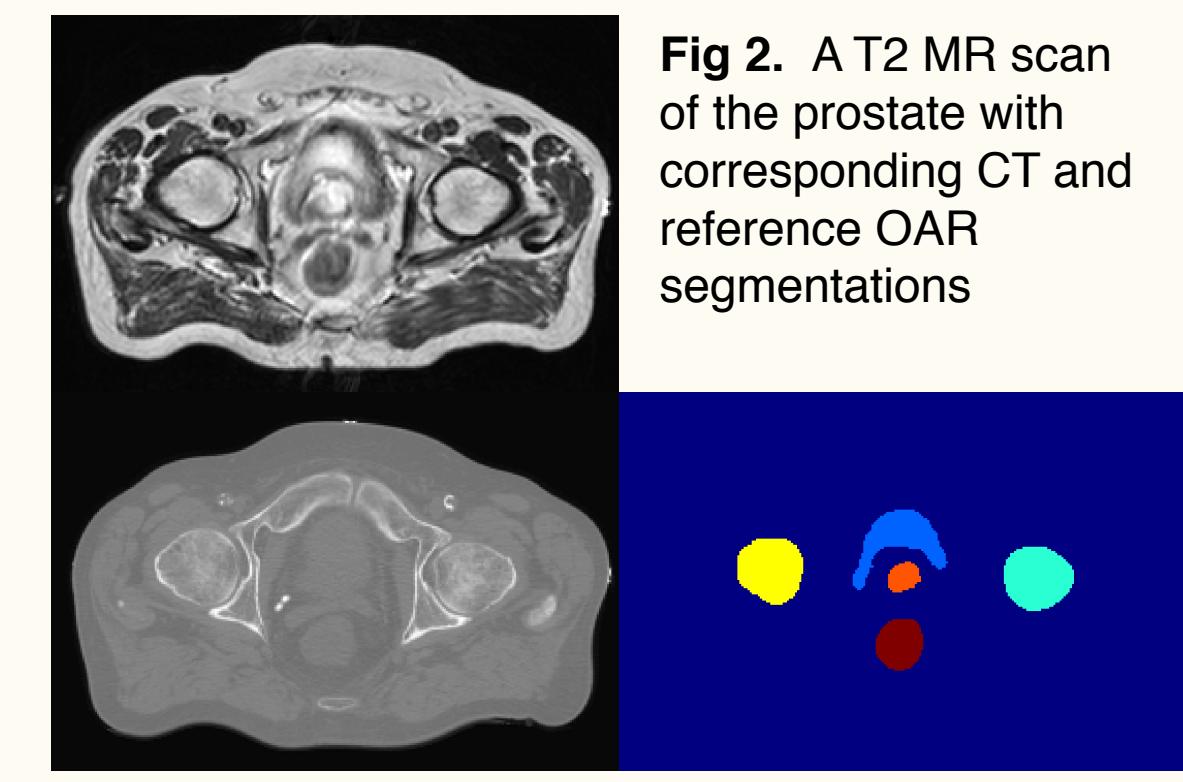
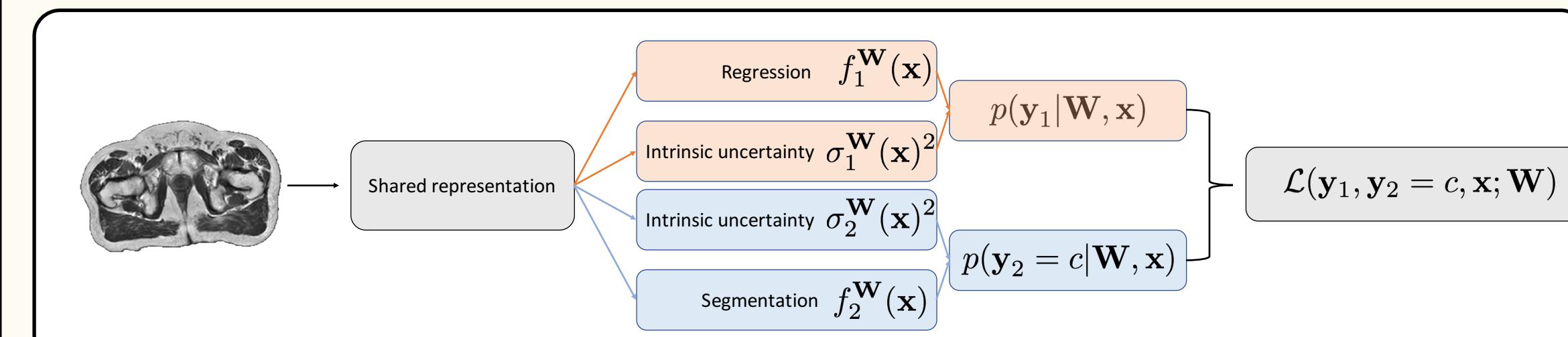
- **Multi-task learning** provides a mechanism that jointly integrates information from distinct sources via inductive transfer [1]
- The quality of inductive transfer is dependent on the relative weighting of task specific losses and the mechanism for sharing network weights. Task loss weightings are hyper-parameters or learned [2]
- We propose a **probabilistic multi-task network** (Fig. 1) that:
  - a) estimates **heteroscedastic uncertainty** for spatially adaptive task loss weighting on a voxel-wise basis
  - b) captures **model uncertainty** through approximate Bayesian inference
- We demonstrate in **MR-only radiotherapy treatment planning**, which requires the synthesis of a CT scan (synCT) from MRI and the segmentation of organs at risk (OAR) (Fig. 2)
- **Results** show:
  1. heteroscedastic uncertainty **improves** multi-task learning over learned task loss weights
  2. the estimated uncertainty can be exploited for **quality control** of the network

## Implementation details



## Probabilistic dual task neural network

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



## Task weighting with heteroscedastic uncertainty

- We extend the modelling of Tanno et al. [4] and Kendall et al. [5] to enrich the homoscedastic multi-task weighting method of Kendall et al. [2]
- Heteroscedastic uncertainty represents the inherent ambiguity present in mapping MR to CT intensity or voxel-wise class memberships

### Regression heteroscedastic noise model

$$\text{Likelihood function} \\ p(y_1|W, x) = \mathcal{N}(f_1^W(x), \sigma_1^W(x)^2)$$

### Segmentation heteroscedastic noise model

$$\text{Likelihood function} \\ p(y_2|W, x) = \text{Softmax}(f_2^W(x)/2\sigma_2^W(x)^2)$$

### Multi-task likelihood

$$\mathcal{L}(y_1, y_2 = c, x; W) = \frac{\|y_1 - f_1^W(x)\|^2}{2\sigma_1^W(x)^2} + \frac{\text{CE}(f_2^W(x), y_2 = c)}{2\sigma_2^W(x)^2} + \log(\sigma_1^W(x)^2 \sigma_2^W(x)^2)$$

## Model uncertainty

- We account for parameter uncertainty through an approximation of the posterior distribution over the network weights
- The posterior distribution is approximated through Bernoulli binary dropout [6]

### Posterior distribution approximation

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

### Sampling posterior during training

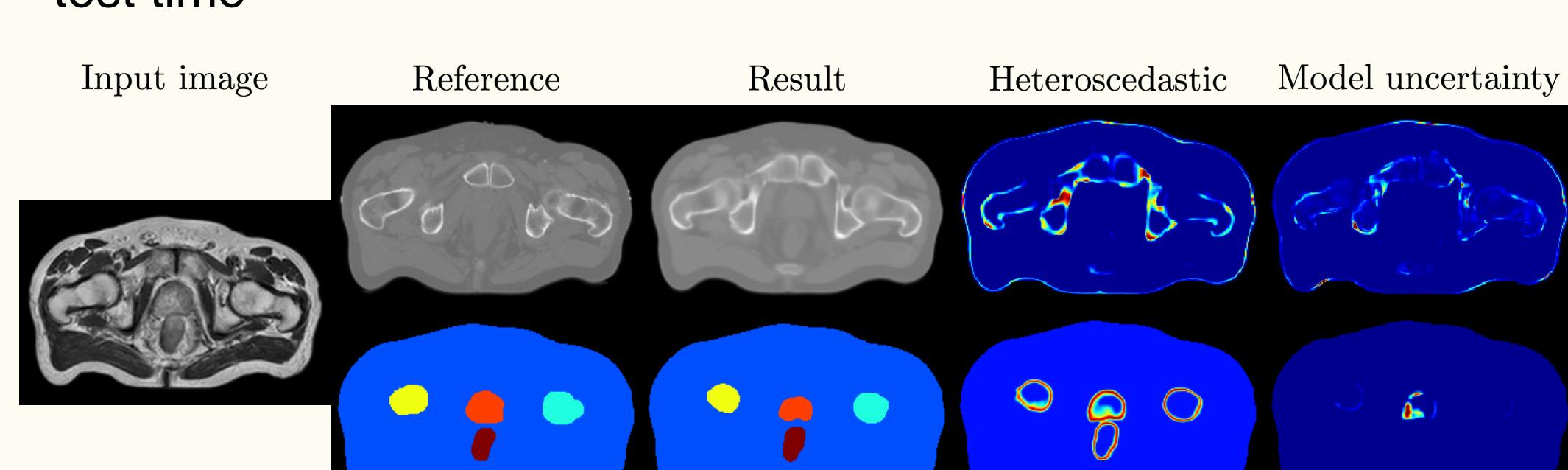
$$w' \sim q(W) \quad f^{w'}(x) := [f_1^{w'}(x), f_2^{w'}(x), \sigma_1^{w'}(x)^2, \sigma_2^{w'}(x)^2]$$

## Quantifying total uncertainty over predictions

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

## Model performance

- We tested on 15 prostate scans with 3-fold cross-validation.
- We trained the network on randomly sampled axial slices and reconstructed the volume at test time

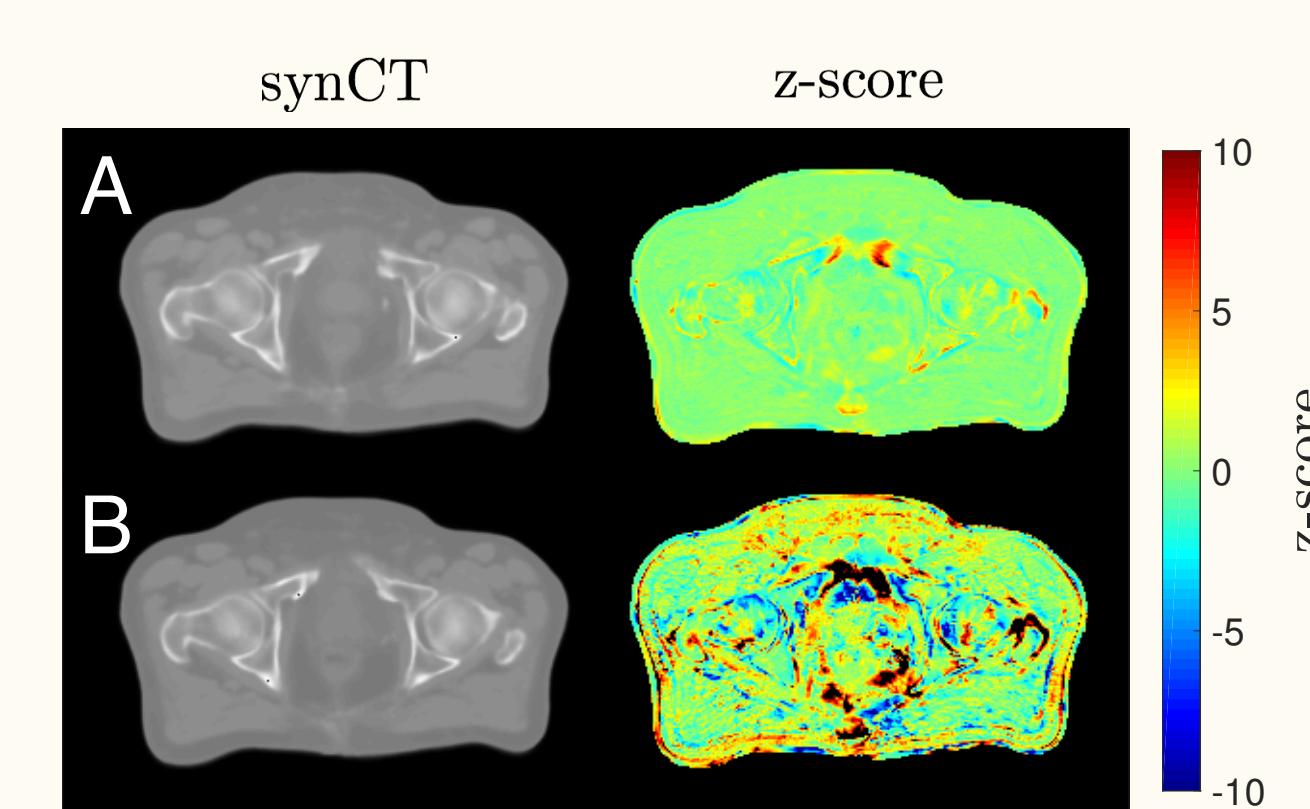
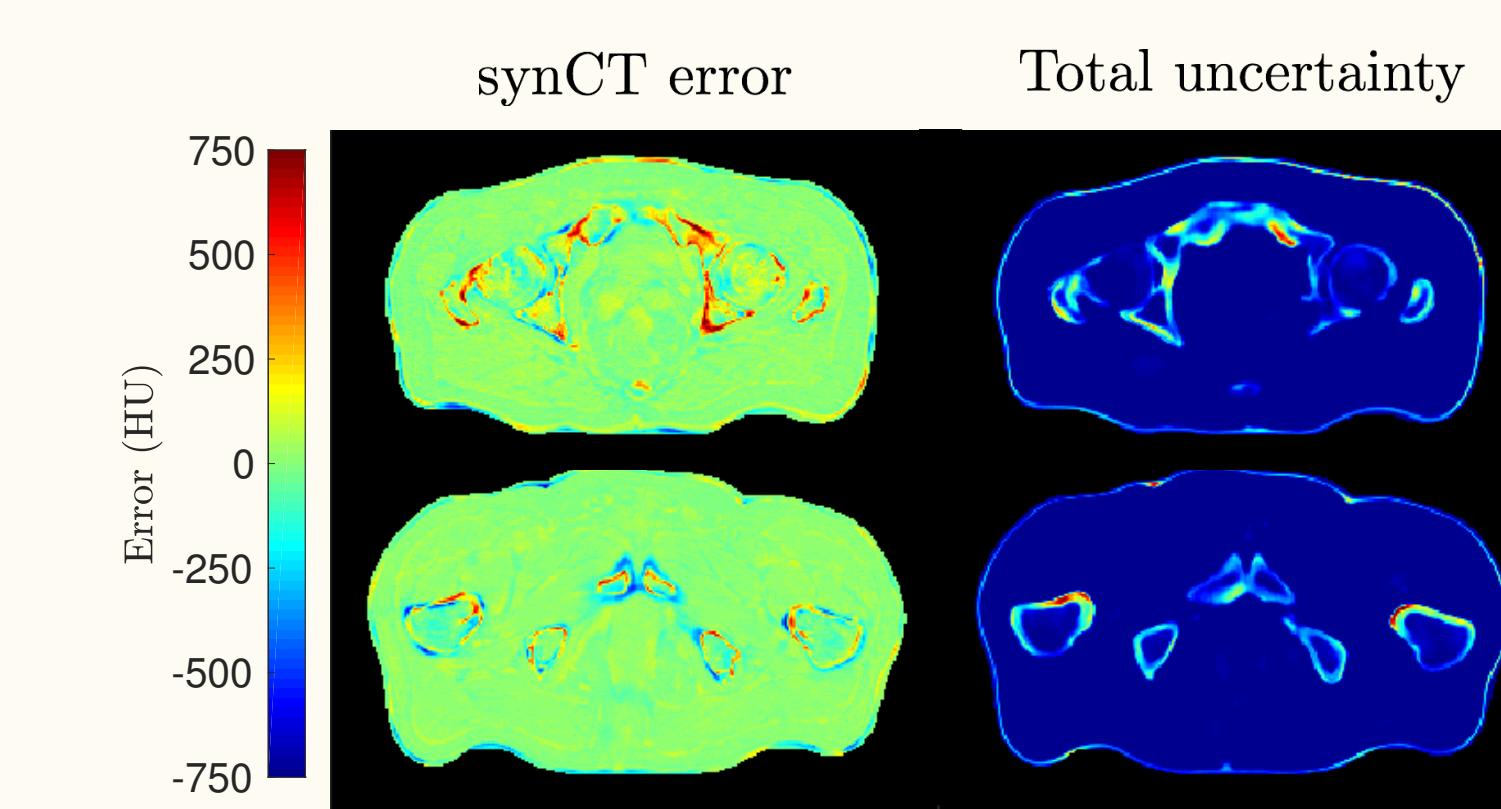


- Joint modelling of heteroscedastic and parameter uncertainty achieves best performance on synCT regression and outperforms homoscedastic task weighting
- Equivalent results in segmentation with the state of the art in pelvic segmentation [7]

Models	All	Bone	L femur	R femur	Prostate	Rectum	Bladder
Regression - synCT - Mean Absolute Error (HU)							
HighResNet [3]	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 [5]	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 [2]	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 [7]	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 [3]	-	-	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(0.02)	0.66(0.12)	0.69(0.13)	0.90(0.07)
HighResNet + dropout + hetero [5]	-	-	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 [2]	-	-	0.92(0.02)	0.92(0.02)	0.73(0.07)	0.76(0.10)	0.93(0.02)
Multi-atlas propagation [7]	-	-	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

- The total uncertainty highly correlates with regression errors – providing a means for automated quality control in absence of a ground truth (Fig. 4)
- Estimation of the total uncertainty (hetero + parameter) provides a better estimation in the variance of the errors (Fig. 5)



- Physiological differences across patients can lead to areas of high uncertainty and errors in the synCT

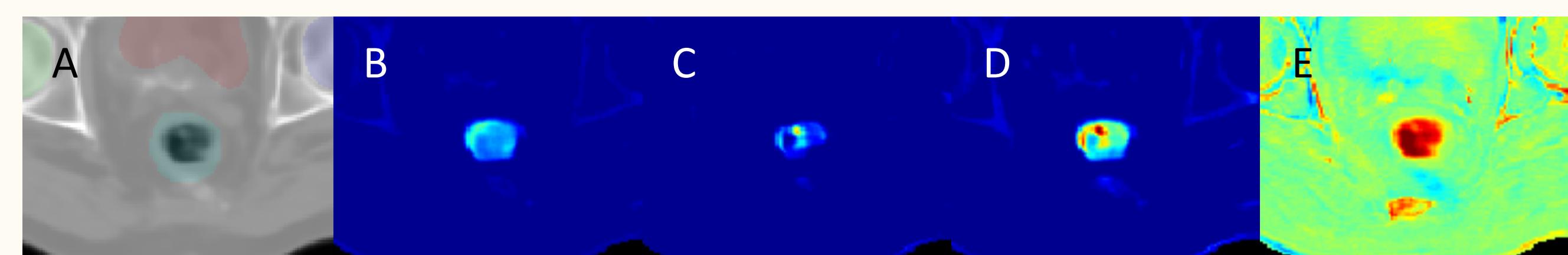


Fig 6. A) Reference segmentation overlaid on MR scan with irregular physiology, B) hetero noise, C) predictive uncertainty, D) total uncertainty and E) synCT error.

## References

1. R. Caruana. In *Learning to learn*, 95-133, 1998
2. A. Kendall et al. In *CVPR*, 2018
3. W. Li et al. In *IPMI*, 348-360, 2017
4. R. Tanno et al. In *MICCAI*, 611-619, 2017
5. A. Kendall et al. In *NIPS*, 5580-5590, 2017
6. Y. Gal and Z. Ghahramani. In *ICML*, 1050-1059, 2016
7. N. Burgos et al. In *Phys. Med. Bio.* 2017

**Acknowledgments.** This work was partially 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.



cmic  
Centre for Medical Image Computing

KING'S COLLEGE LONDON  
National University of Singapore

CANCER RESEARCH UK  
ART-NET

