Bayesian Image Quality Transfer with CNNs: Exploring Uncertainty in dMRI Super-Resolution

Ryutaro Tanno[1], Daniel E. Worrall[1], Aurobrata Ghosh[1], Enrico Kaden[1], Stamatios N. Sotiropoulos[2], Antonio Criminisi[3], Daniel C. Alexander[1]

[1] Dept. of Computer Science, University College London, UK
[2] FMRIB Centre, University of Oxford, Oxford, UK
[3] Machine Intelligence and Perception Group, Microsoft Research Cambridge, UK

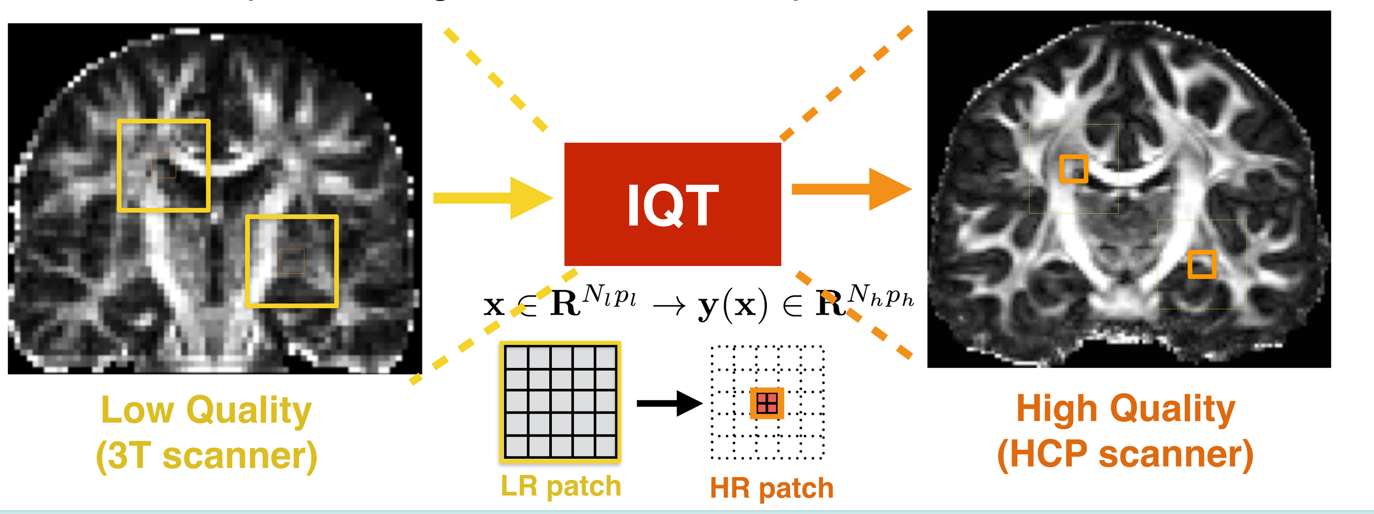
UCL

Abstract

- **Image quality transfer (IQT)** [1] is a machine-learning based framework to enhance low quality images (e.g. clinical data) by learning and propagating rich information from rare high quality images from expensive scanners (e.g. HCP data).
- We propose a **Bayesian extension of IQT** based on **probabilistic deep learning** methods.
- We demonstrate in **super-resolution of dMRI.**
- ***Results*** show:
    1. our method **improves reconstruction accuracy.**
    2. our method shows tangible **benefits in downstream tractography.**
    3. our method provides a means to estimate **uncertainty over prediction,** which can be used as a surrogate measure of accuracy.

Background (IQT framework)

- **Super-resolution as a patch-wise regression** as in [1, 2].
- **Training data generation**: high quality images from HCP are downsampled to create matched pairs of high-res and low-res patches.

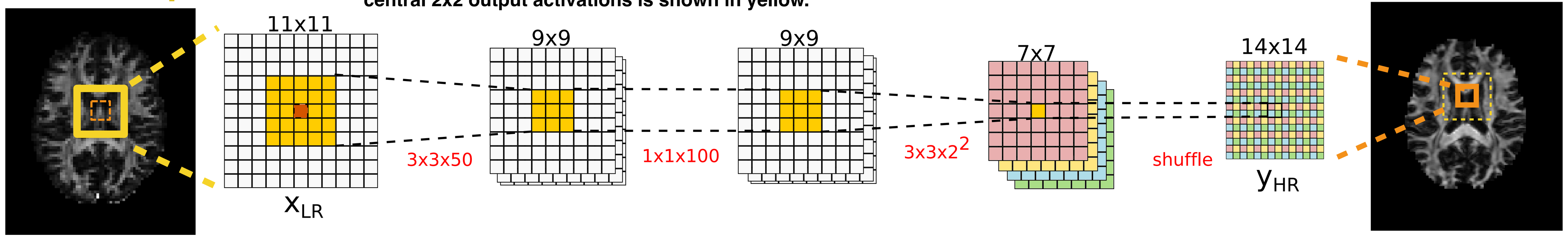


Methods

1. *Baseline 3D Super-resolution Network*

- **3D Extension of ESPCN [3]**
- Minimal architecture (3 conv. + shuffle)
- Trained to minimise pixel-wise MSE

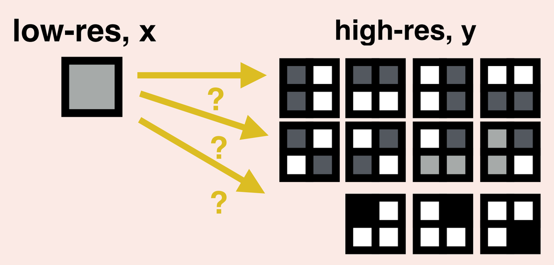
Fig.1. 2D illustration of the baseline network with upsampling rate, r = 2. The receptive field of the central 2x2 output activations is shown in yellow.



2. *Probabilistic CNNs: model **two** types of uncertainty*

Type (I): Intrinsic uncertainty

- inherent ambiguity in the problem
- irreducible even with **infinite** data.



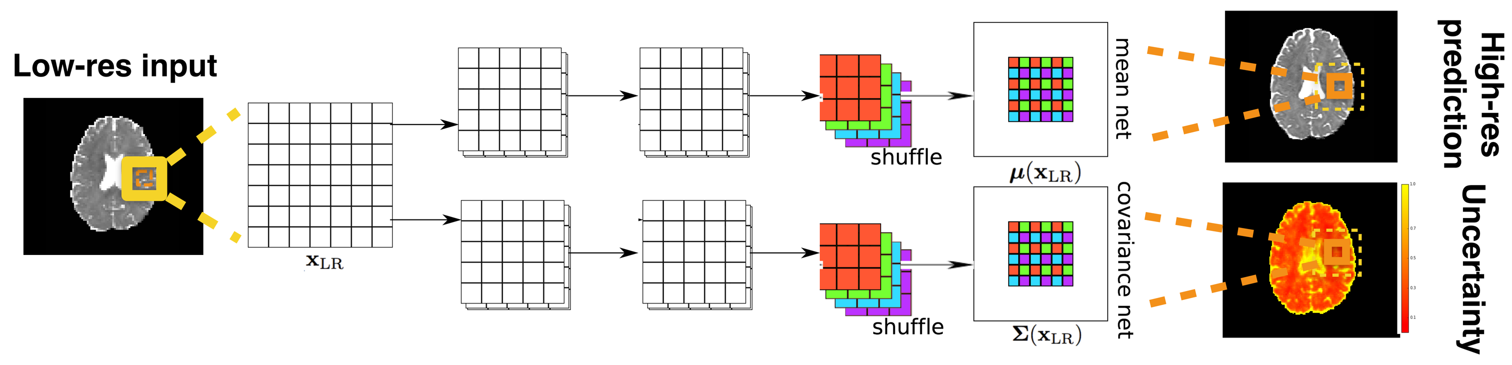
Solution (I): Heteroscedastic noise model

- model super-resolution mapping as a **spatially varying** multivariate Gaussian dist [4]

$$p(\mathbf{y}|\mathbf{x}, \theta, \mathcal{D}) = \mathcal{N}(\mathbf{y}; \mu_\theta(\mathbf{x}), \Sigma_\theta(\mathbf{x}))$$

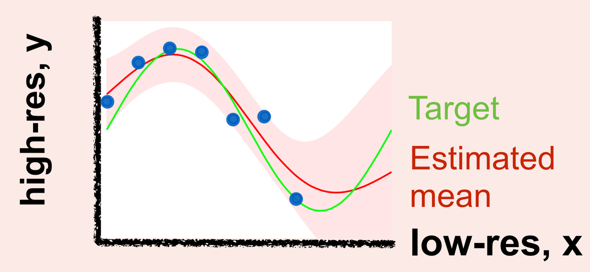
- **dual architecture:** use two separate 3D-ESPCNs to model the mean and the covariance (see Fig. 2).
- **diagonal components** in the covariance estimates intrinsic uncertainty

Fig.2. the dual architecture for estimating intrinsic uncertainty. Diagonal covariance is assumed. The top 3D-ESPCN estimates the mean and the bottom one estimates the covariance matrix of the likelihood. The diagonal entries of the covariance matrix estimates the intrinsic uncertainty.



Type (II): Parameter uncertainty

- ambiguity in the "best" model
- can be explained away by **infinite** data.



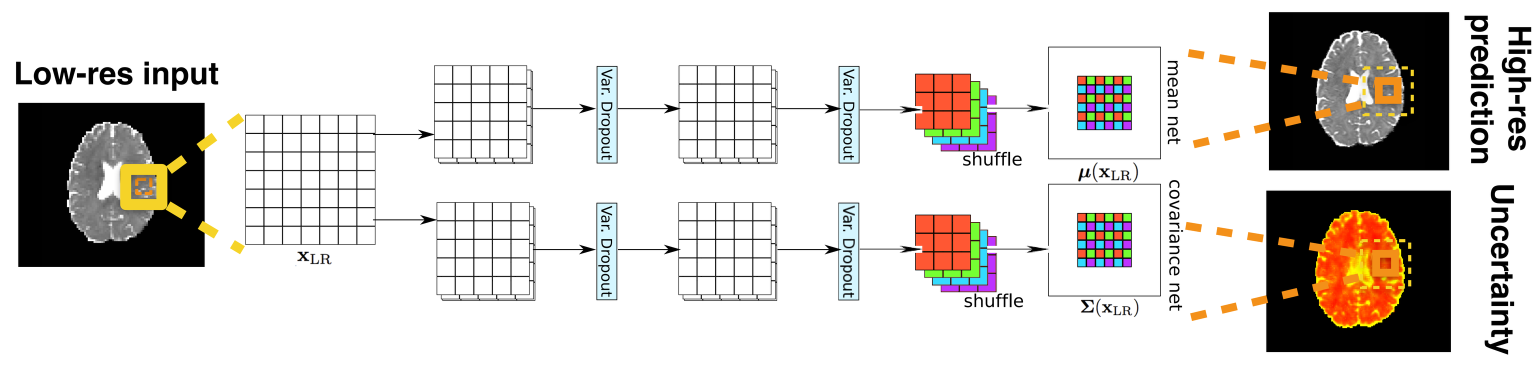
Solution (II): Variational drop-out

- **Average over all possible models** weighted by the posterior over the weights i.e

$$\underbrace{p(\mathbf{y}|\mathbf{x}, \mathcal{D})}_{\text{predictive dist.}} = \int \underbrace{\mathcal{N}(\mathbf{y}; \mu_\theta(\mathbf{x}), \Sigma_\theta(\mathbf{x}))}_{\text{hetero. likelihood}} \cdot \underbrace{p(\theta|\mathbf{x}, \mathcal{D})}_{\text{posterior dist.}} \, d\theta$$

- Approximate **posterior** with a Gaussian dist. using variational drop-out [5]
- At test time, the "learned" Gaussian **noise is injected** into every convolutional filter (Fig. 3).

Fig.3. combine variational dropout with the heteroscedatisc network to model parameter uncertainty. After every convolution, "learned" noise is injected into the feature maps via the "variational dropout layer".



3. *Quantifying uncertainty over prediction:*

at test time, *given an input patch **x**, apply MC dropout - run multiple forward passes, inject noise according to the likelihood and collect many samples of high-res outputs* $\{\mathbf{y}^{(1)}, \mathbf{y}^{(2)}, ..., \mathbf{y}^{(T)}\}$. *Then, estimate the mean **(predictive mean)** and standard deviation* **(predictive uncertainty)**. Use predictive mean as the **final estimate of y** and predictive uncertainty to quantify its **confidence**.

Results

1. x2 DTI super-resolution

- Evaluated performance on **two** datasets.
- The **baseline CNN (3D-ESPCN)** outperforms the current state-of-the-art model (BIQT-Random-Forests.).
- **An order of magnitude faster**: 1s on a GPU and 10s on a CPU.
- Jointly modelling intrinsic uncertainty (Hetero-Noise) and parameter uncertainty (Variational-Dropout) achieves the best performance.

best | 2nd best

Models	HCP (rmse)	Lifespan (rmse)
Cubic Interpolation	10.069± n/a	32.483± n/a
β-Spline Interpolation	9.578± n/a	33.429± n/a
IQT-Random-Forests	6.974 ± 0.024	10.038 ± 0.019
BIQT-Random-Forests	6.972 ± 0.069	9.926 ± 0.055
3D-ESPCN(baseline)	6.378 ± 0.015	8.998 ± 0.021
Binary-Dropout-CNN($p = 0.1$)	6.963 ± 0.034	9.784 ± 0.048
Gaussian-Dropout-CNN($p = 0.1$)	6.519 ± 0.015	9.183 ± 0.024
Variational-Dropout(I)-CNN	6.354 ± 0.015	8.973 ± 0.024
Variational-Dropout(II)-CNN	6.356 ± 0.008	8.982 ± 0.024
Hetero-Noise-CNN	6.294 ± 0.029	8.985 ± 0.051
Hetero-Noise+Variational-Dropout(I)	6.291 ± 0.012	$\mathbf{8.944 \pm 0.044}$
Hetero-Noise+Variational-Dropout(II)	$\mathbf{6.287 \pm 0.029}$	8.955 ± 0.029

2. Benefits in downstream processing: tractography

- (yellow arrows): **CNN avoids a false positive** tract better than RF and Linear Interp.
- (green arrows): **CNN achieves shaper recovery of WM tracts.**

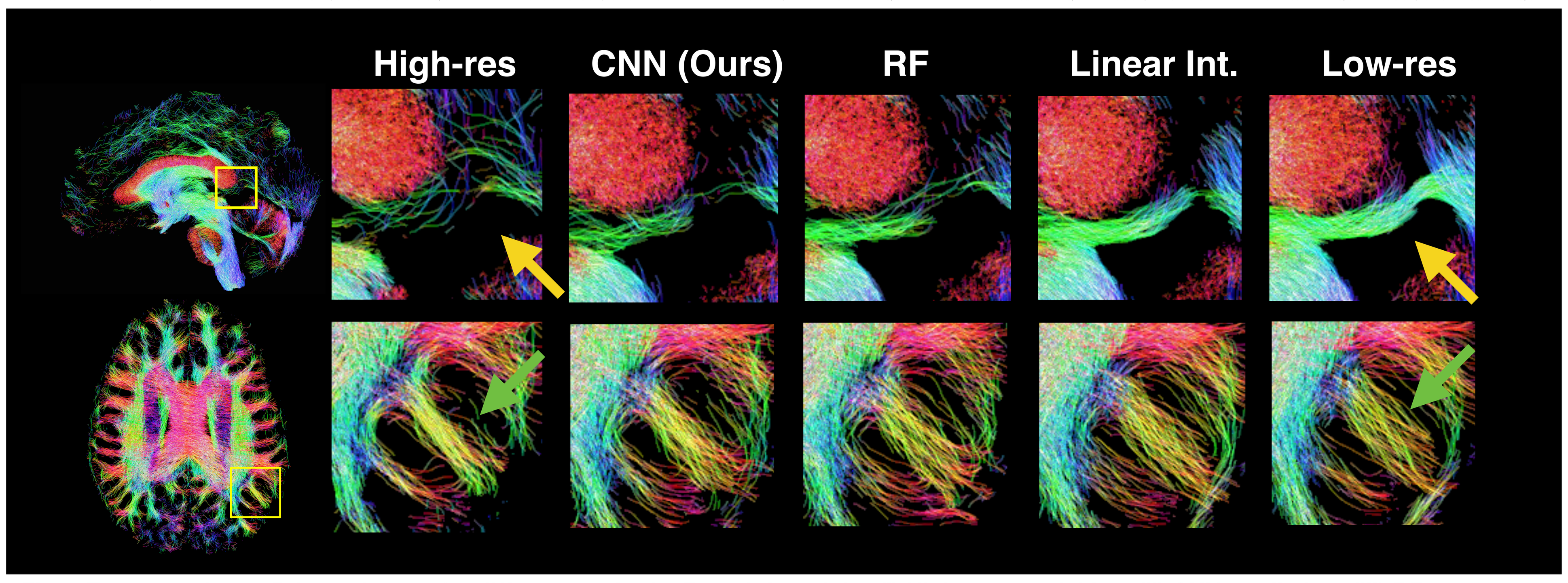


Fig.4. Tractography on Prisma dataset for different methods. From left to right: (i) High-res acquisition, (ii) CNN prediction; (iii) RF; (iv) Linear interpolation; (v) Low-res acquisition.

3.Visualisation of predictive uncertainty

- predictive mean and uncertainty are estimated from 200 samples of high-res DTIs.
- **high correlation** between the uncertainty map and error map (Fig. 5)
- **highlight pathology** not represented in the training data (Fig. 6)

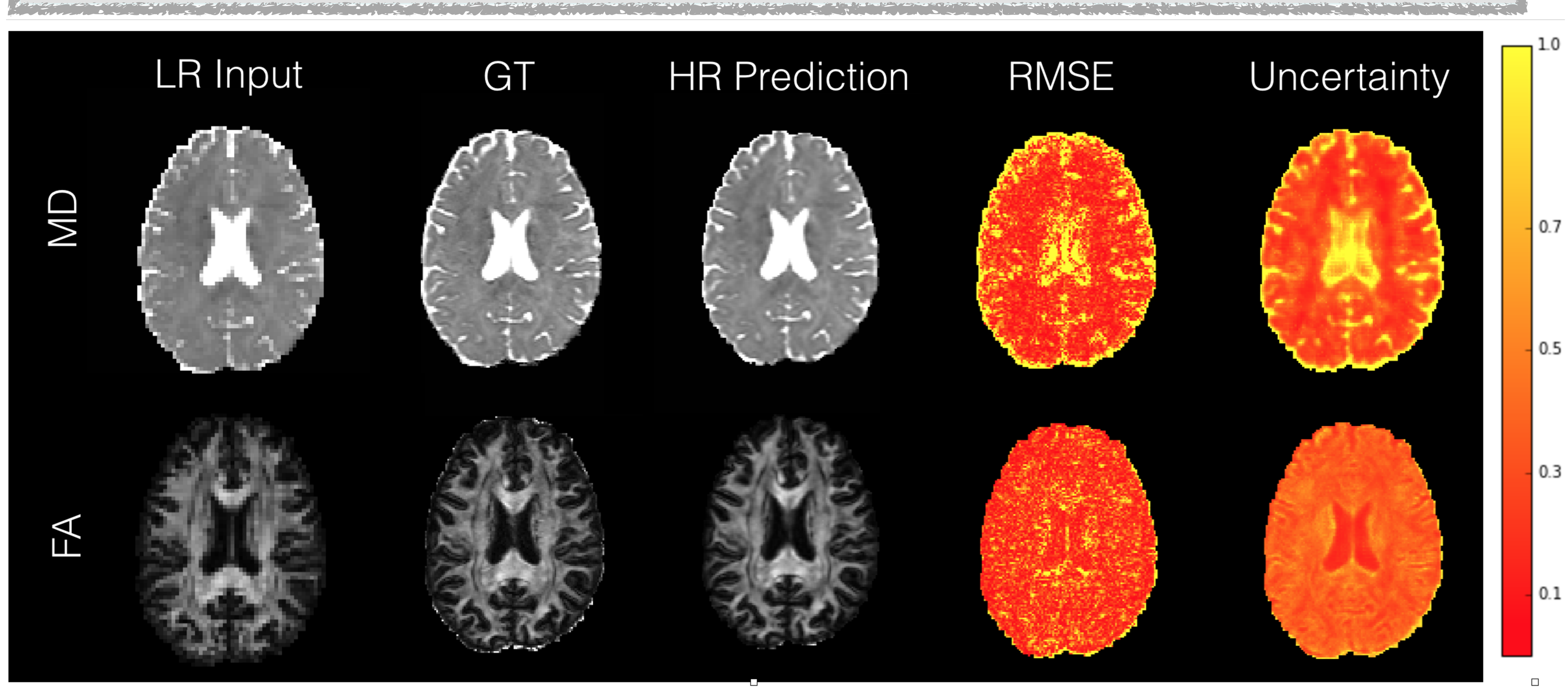


Fig.5 Comparison between RMSE and uncertainty maps for FA and MD computed on a HCP subject. LR input, ground truth and HR prediction are also shown.

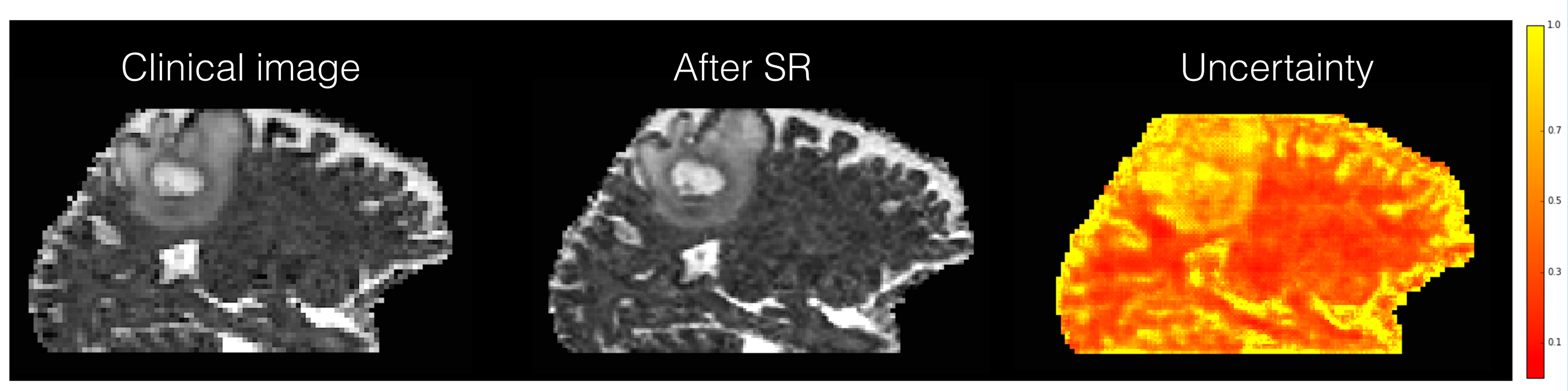


Fig.6 DTI SR on a brain tumour patient. From top to bottom: (i) MD computed from the original DTI; (ii) the estimated HR version; (iii) uncertainty.

References:
1. Alexander, D.C., et al.: Image quality transfer via random forest regression. In: Neuroimage 2017.
2. Tanno, R, et al.: Bayesian image quality transfer. In: MICCAI 2016
3. Shi, W., et al.: Real-time single image and video super-resolution using an efficient sub-pixel convolutional neural network, CVPR 2016
4. Nix, D, et al: Estimating the mean and variance of the target probability distribution, 1994
5. Kingma, D, et al: Variational dropout and the local reparametrization trick, NIPS 2015

Microsoft Research Cambridge
cmic Centre for Medical Image Computing
NHS National Institute for Health Research