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

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

[1] Department of Computer Science, University College London, UK,
[2] FMRIB Centre, University of Oxford, UK,
[3] Microsoft Research Cambridge, UK

**Abstract.** In this work, we investigate the value of uncertainty modelling in 3D super-resolution with convolutional neural networks (CNNs). Deep learning has shown success in a plethora of medical image transformation problems, such as super-resolution (SR) and image synthesis. However, the highly ill-posed nature of such problems results in inevitable ambiguity in the learning of networks. We propose to account for *intrinsic uncertainty* through a per-patch heteroscedastic noise model and for *parameter uncertainty* through approximate Bayesian inference in the form of variational dropout. We show that the combined benefits of both lead to the state-of-the-art performance SR of diffusion MR brain images in terms of errors compared to ground truth. We further show that the reduced error scores produce tangible benefits in downstream tractography. In addition, the probabilistic nature of the methods naturally confers a mechanism to quantify uncertainty over the super-resolved output. We demonstrate through experiments on both healthy and pathological brains the potential utility of such an uncertainty measure in the risk assessment of the super-resolved images for subsequent clinical use.

## 1 Introduction and Background

Algorithmic and hardware advancements of non-invasive imaging techniques, such as MRI, continue to push the envelope of quality and diversity of obtainable information of the underlying anatomy. However, their prohibitive cost and lengthy acquisition time often hinder the translation of such technological innovations into clinical practice. Poor image quality limits the accuracy of subsequent analysis, potentially leading to false clinical conclusions. Therefore, methods which can efficiently and reliably boost scan quality are in demand.

Numerous machine learning based methods have been proposed for various forms of image enhancement, generally via supervised regression of low quality (e.g., clinical) against high quality (e.g., experimental) image content. Alexander et al. [1] propose a general framework for supervised quality enhancement, which they call image quality transfer (IQT). They demonstrated this with a random forest (RF) implementation of super-resolution (SR) of brain diffusion tensor images (DTIs) and estimation of advanced microstructure parameter maps from

sparse measurements. More recently, deep learning has shown additional promise in this kind of task. For example, [2] proposed a CNN model to upsample a stack of 2D MRI cardiac volumes in the through-plane direction. Another application of CNNs is the prediction of 7T images from 3T MRI [3], where both contrast and resolution are enhanced. Current methods typically commit to a single prediction, leaving users with no measure of prediction reliability. One exception is Bayesian IQT [4], which proposes a variant of RF to quantify predictive uncertainty over high-resolution (HR) DTIs and demonstrate its utility as a surrogate measure of accuracy.

This paper proposes a new implementation of Bayesian IQT via CNNs. This involves two key innovations in CNN-based models: 1) we extend the subpixel CNN of [5], previously limited to 2D images, to 3D volumes, outperforming previous models in accuracy and speed on a DTI SR task; 2) we devise new architectures enabling estimates of different components of the uncertainty in the SR mapping. The first enables us to bring the performance benefits of deep learning to this important problem, as well as reducing computation time to super-resolve the entire brain DTI in 1 s. For our second contribution, we describe two kinds of uncertainty which arise when tackling image enhancement problems. The first kind of uncertainty, which we call *intrinsic uncertainty* is defined as the irreducible variance of the statistical mapping from low-resolution (LR) to HR. This inherent ambiguity arises from the fact that the LR to HR problem is one-to-many, and is present independent of the amount of data we collect. We model the variation in intrinsic uncertainty over different structures within the anatomy through a per-patch heteroscedastic noise model [6]. The second kind of uncertainty, which we call *parameter uncertainty*, quantifies the degree of ambiguity in the model parameters that best explain the observed data, which arises from the finite training set. We account for it through approximate Bayesian inference in the form of variational dropout [7].

We first evaluate the performance of the proposed probabilistic CNN methods and the benefits of uncertainty modelling by measuring the deviation from the ground truth on standard metrics. Human Connectome Project (HCP) dataset [8] and the Lifespan dataset (http://lifespan.humanconnectome.org/) are used for the quantitative comparison. We also test its practical benefits in downstream tractography through SR of Mean Apparent Propagator (MAP)-MRI [9]. Lastly, we investigate the utility of uncertainty maps over the predicted HR images by testing on images of both healthy subjects and brain tumour patients.

## 2 Method

As in [1–3], we formulate the SR task as a patch-wise regression where an input LR image is split into smaller overlapping sub-volumes and the resolution of each is sequentially enhanced. We first propose a baseline model. We then build on it by integrating two complementary methods for assimilating uncertainty.

**Baseline network:** Efficient subpixel-shifted convolutional network (ESPCN) [5] is a recently proposed method with the capacity to perform real-time per-

frame SR of videos while retaining cutting-edge performance. We extend this method to 3D and use this as our baseline model (3D-ESPCN). Most CNN-based SR techniques [2, 10, 11] first up-sample a low-resolution (LR) input image (e.g. through bilinear interpolation, deconvolution, fractional-strided convolution, etc.) and then refine the high-resolution (HR) estimate through a series of convolutions. These methods suffer from the fact that (1) the up-sampling can be a lossy process and (2) refinement in the HR-space has a higher computational cost than in the LR-space. ESPCN performs convolutions in the LR-space, upsampling afterwards. The reduced resolution of feature maps dramatically decreases the computational and memory costs, which is more pronounced in 3D.

More specifically the ESPCN is a fully convolutional network, with a special *shuffling operation* on the output (see Fig. 1). The fully convolutional part of the network consists of 3 convolutional layers, each followed by a ReLU, where the final layer has $cr^2$ channels, $r$ being the upsampling rate. The shuffling operation takes an input of shape $h \times w \times cr^2$ and remaps pixels from different channels into different spatial locations in the HR output, producing a $rh \times rw \times c$ image, where $h$, $w$ and $c$ denote height, width and number of channels. This shufflling operation in 3D is $\mathcal{S}(F)_{i,j,k,c} = F_{[i/r],[j/r],[k/r],(r^3-1)c+\mathrm{mod}(i,r)+r\cdot\mathrm{mod}(j,r)+r^3\cdot\mathrm{mod}(k,r)}$ where $F$ is the pre-shuffled feature maps, and is equivalent to learned interpolation.

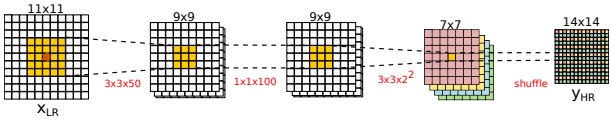


**Fig. 1:** 2D illustration of the baseline network with upsampling rate, $r = 2$. The receptive field of the central $2^2$ output activations is shown in yellow.

At test time, the network takes each sub-volume $\mathbf{x}$ in a LR image, and predicts the corresponding HR sub-volume $\mathbf{y}$. The network increases the resolution of the central voxel of each receptive field, e.g. the central $2^3$ output voxels are estimated from the corresponding $5^3$ receptive field in the input, coloured yellow in Fig.1. Tessellating predictions from shifted $\mathbf{x}$ recovers the whole HR volume.

Given a training set $\mathcal{D} = \{(\mathbf{x}_i, \mathbf{y}_i)\}_{i=1}^N$, we optimize the network parameters by minimising the sum of per-pixel mean-squared-error (MSE) between the ground truth $\mathbf{y}$ and the predicted HR patch $\mu_\theta(\mathbf{x})$ over the training set. $\theta$ denotes all network parameters. This is equivalent to minimising the negative log likelihood (NLL) under the Gaussian noise model $p(\mathbf{y}|\mathbf{x}, \theta) = \mathcal{N}(\mathbf{y}; \mu_\theta(\mathbf{x}), \sigma^2 I)$. Here, HR patches are modelled as a function of LR patches corrupted by isotropic noise with variance $\sigma^2$. Assuming that the model is correct, the variance $\sigma^2$ signifies the degree of irreducible uncertainty in the prediction of $\mathbf{y}$ given $\mathbf{x}$, and thus the *intrinsic uncertainty* in the SR mapping defined in the introduction. However, the quality of this intrinsic uncertainty estimate is limited by the likelihood model; the baseline network assumes constant uncertainty across all spatial locations and image channels, which is over-simplistic for most medical images.

**Heteroscedastic likelihood:** We introduce a *heteroscedastic* noise model to approximate the variation in intrinsic uncertainty across the image. The likelihood becomes $p(\mathbf{y}|\mathbf{x}, \theta_1, \theta_2) = \mathcal{N}(\mathbf{y}; \mu_{\theta_1}(\mathbf{x}), \Sigma_{\theta_2}(\mathbf{x}))$ where both mean and covariance are estimated by two separate 3D-ESPCNs $\mu_{\theta_1}(\cdot)$ and $\Sigma_{\theta_2}(\cdot)$ as functions of the input. The mean network makes predictions and the covariance network estimates the intrinsic uncertainty (see Fig. 2). The diagonal of $\Sigma_{\theta_2}(\mathbf{x})$ quantify the estimated intrinsic uncertainty over individual components in $\mu_{\theta_1}(\mathbf{x})$.

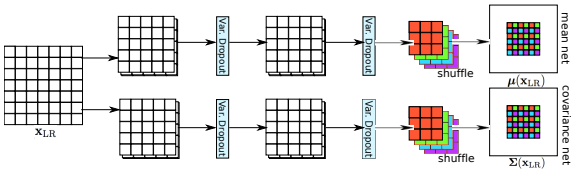


**Fig. 2:** 2D illustration of a heteroscedastic network with variational dropout. Diagonal covariance is assumed. The top 3D-ESPCN estimates the mean and the bottom one estimates the covariance matrix of the likelihood. Variational dropout is applied to feature maps after every convolution.

The NLL is $\mathcal{L}_\theta(\mathcal{D}) = \mathcal{H}_\theta(\mathcal{D}) + \mathcal{M}_\theta(\mathcal{D})$ with $\mathcal{H}_\theta(\mathcal{D}) = \frac{1}{N}\sum_{i=1}^{N} \log \det \Sigma_{\theta_2}(\mathbf{x}_i)$ i.e. mean differential entropy and $\mathcal{M}_\theta(\mathcal{D}) = \frac{1}{N}\sum_{i=1}^{N} (\mathbf{y}_i - \mu_{\theta_1}(\mathbf{x}_i))^T \Sigma_{\theta_2}^{-1}(\mathbf{x}_i)(\mathbf{y}_i - \mu_{\theta_1}(\mathbf{x}_i))$ i.e. mean squared Mahalanobis distance. $\mathcal{M}_\theta(\mathcal{D})$ seeks to minimise the weighted MSE under the covariance while $\mathcal{H}_\theta(\mathcal{D})$ controls the 'spread' of $\Sigma_{\theta_2}(\mathbf{x})$.
**Bayesian inference through variational dropout:** The baseline 3D-ESPCN and heteroscedastic model neglect *parameter uncertainty*, relying on a single estimate of the network parameters. In medical imaging where data size is commonly limited, this point-estimate approach potentially leads to overfitting. We combat this with a Bayesian approach, averaging over all possible models $p(\mathbf{y}|\mathbf{x}, \theta)$ weighted by the (posterior) probability of the parameters given the training data, $p(\theta|\mathcal{D})$ i.e. we aim to compute $p(\mathbf{y}|\mathbf{x}, \mathcal{D}) = \mathbb{E}_{p(\theta|\mathcal{D})}[p(\mathbf{y}|\mathbf{x}, \theta)]$. However, this expectation is intractable. Variational dropout [7] addresses this problem for neural networks, using a form of variational inference where the posterior $p(\theta|\mathcal{D})$ is approximated by a factored Gaussian distribution $q_\phi(\theta) = \prod_{ij} \mathcal{N}(\theta_{ij}; m_{ij}, s_{ij}^2)$. During training, the network learns the parameters $\phi = \{m_{ij}, s_{ij}^2\}$.

At test time, given a LR input $\mathbf{x}$, we estimate the mean and covariance of the approximate predictive distribution $q_\phi^*(\mathbf{y}|\mathbf{x}) \triangleq \mathbb{E}_{q_\phi(\theta)}[p(\mathbf{y}|\mathbf{x}, \theta)]$ with the MC estimators $\hat{\mu}_{\mathbf{y}|\mathbf{x}} \triangleq \frac{1}{T}\sum_{t=1}^{T} \mu_{\theta_1^t}(\mathbf{x})$ and $\hat{\Sigma}_{\mathbf{y}|\mathbf{x}} \triangleq \frac{1}{T}\sum_{t=1}^{T} (\Sigma_{\theta_2^t}(\mathbf{x}) + \mu_{\theta_1^t}(\mathbf{x})\mu_{\theta_1^t}(\mathbf{x})^T) - \hat{\mu}_{\mathbf{y}|\mathbf{x}}\hat{\mu}_{\mathbf{y}|\mathbf{x}}^T$, where $\theta^t = (\theta_1^t, \theta_2^t)$ are samples of weights from the approximate posterior $q_\phi(\theta)$. We use the sample mean as the final prediction of an HR patch and the diagonal of the sample variance as the corresponding uncertainty. When we use the baseline model, the first term in the sample variance reduces to $\sigma^2 I$.
**Implementation details:** We employed a common protocol for the training of all networks. We minimized the loss using ADAM [12] for 200 epochs

with learning rate $10^{-3}$. As in [5], we use a minimal architecture for the baseline 3D-ESPCN, consisting of 3 convolutional layers with filters $(3,3,3,50) \rightarrow (1,1,1,100) \rightarrow (3,3,3,r^3c)$ where $r$ is upsampling rate and $c$ is the number of channels. The filter sizes are chosen so a $(5,5,5)$ LR patch maps to a $(r,r,r)$ HR patch, which mirrors random forest based methods [1, 4] for a fair comparison. The heteroscedastic network of Section 2 is formed of two 3D-ESPCNs, separately estimating the mean and standard deviations. Positivity of the standard deviations is enforced by passing the output through a *softplus* function. For variational dropout we tried two flavours: Var.(I) optimises per-weight dropout rates, and Var.(II) optimises per-filter dropout rates. Variational dropout is applied to both the baseline and heteroscedastic models without changing the architectures.

All models are trained on datasets generated from 8 randomly selected HCP subjects [8], each consisting of 90 diffusion weighted images (DWIs) of voxel size $1.25^3$ mm$^3$ with $b = 1000$ s/mm$^2$. The training set is created by sampling HR subvolumes from the ground truth DTIs (or MAP-MRI coefficients) and then downsampling to generate the LR counterparts. Downsampling is done in the raw DWI by a factor of $r$ by taking a block-wise mean and then the DT or MAP coefficients are subsequently computed. Each network is trained on $\sim 4000$ pairs of input/output patches of size $11^3c$ and $(7r)^3c$, amounting to $\sim 1.4 \times 10^6$ receptive field patch pairs of dimensions $5^3c$ and $r^3c$, which is roughly the same size as the maximal training set used in RF-IQT [**?**]. It takes under 30/120 mins to train a single network on DTI/MAP-MRI data on 1 TITAN X GPU.

## 3 Experiments and Results

**Performance comparison for DTI SR:** We evaluate the performance of our models for DTI SR on two datasets. The first contains 8 unseen subjects from the same HCP cohort used for training. The second consists of 10 subjects from the HCP Lifespan dataset. The latter tests generalisability, as they are acquired with different protocols, at lower resolution (1.5 mm isotropic), and on subjects of older age range (45-75) to the original HCP data (22-36). We perform $\times 2$ upsampling in each direction, measuring reconstruction accuracy with RMSE, PSNR and MSSIM on interior and exterior separately as shown in Fig. 3(b). This is important, as the estimation problem is quite different in boundary regions, but remains valuable for applications like tractography where seed or target regions are often in the cortical surface of the brain. We only present the RMSE results, but the derived conclusions remain the same for the other two metrics.

Fig. 3(a) shows our baseline achieves 8.5%/39.8% reduction in RMSE on the HCP dataset on interior/exterior regions with respect to the best published method, BIQT-RF[4]. Note that IQT-RF and BIQT-RF are only trained on interior patches, and SR on boundary patches requires a separate ad-hoc procedure. Despite including exterior patches in training our model, which complicates the learning, the baseline CNN out-performs the RF methods on both regions—this goes for the Lifespan dataset too. The 3D-ESPCN estimates whole HR volumes $<$10 s on a CPU and 1 s on a GPU, while BIQT-RF takes 10 mins with 8 trees.

(a) Performance comparison

| Models | HCP (interior) | HCP (exterior) | Life (interior) | Life (exterior) |
|---|---|---|---|---|
| CSpline | 10.069± n/a | 31.738± n/a | 32.483± n/a | 49.066± n/a |
| $\beta$-Spline | 9.578± n/a | 98.169± n/a | 33.429± n/a | 186.049± n/a |
| IQT-RF | $6.974 \pm 0.024$ | $23.139 \pm 0.351$ | $10.038 \pm 0.019$ | $25.166 \pm 0.328$ |
| BIQT-RF | $6.972 \pm 0.069$ | $23.110 \pm 0.362$ | $9.926 \pm 0.055$ | $25.208 \pm 0.290$ |
| 3D-ESPCN(baseline) | $6.378 \pm 0.015$ | $13.909 \pm 0.071$ | $8.998 \pm 0.021$ | $16.779 \pm 0.109$ |
| Dropout-CNN(0.1) | $6.963 \pm 0.034$ | $14.568 \pm 0.068$ | $9.784 \pm 0.048$ | $17.357 \pm 0.091$ |
| Gaussian-CNN(0.1) | $6.519 \pm 0.015$ | $14.038 \pm 0.038$ | $9.183 \pm 0.024$ | $16.890 \pm 0.097$ |
| Var.(I)-CNN | $6.354 \pm 0.015$ | $\mathbf{13.824 \pm 0.031}$ | $8.973 \pm 0.024$ | $\mathbf{16.633 \pm 0.053}$ |
| Var.(II)-CNN | $6.356 \pm 0.008$ | $13.846 \pm 0.017$ | $8.982 \pm 0.024$ | $16.738 \pm 0.073$ |
| Hetero-CNN | $6.294 \pm 0.029$ | $15.569 \pm 0.273$ | $8.985 \pm 0.051$ | $17.716 \pm 0.277$ |
| Hetero+Var.(I) | $6.291 \pm 0.012$ | $13.906 \pm 0.048$ | $\mathbf{8.944 \pm 0.044}$ | $16.761 \pm 0.047$ |
| Hetero+Var.(II) | $\mathbf{6.287 \pm 0.029}$ | $13.927 \pm 0.093$ | $8.955 \pm 0.029$ | $16.844 \pm 0.109$ |

(b) Mask

**Fig. 3:** (a) RMSE on HCP and Lifespan dataset for different upsampling methods. For each, an ensemble of 10 models are trained on different training sets, and the mean/std of the average errors over 8 test subjects are computed over the ensemble. Best results in bold red, and the second best in blue. (b) Interior (yellow) and exterior region (red).

Hetero-CNN improves on the performance of 3D-ESPCN with statistical significance ($p < 10^{-3}$) on the interior region for both HCP and Lifespan data. However, poorer performance is observed on the exterior than the baseline. Using 200 weight samples, Var.(I)-CNN performs best on both datasets on the exterior region. Combination of hetero-CNN and variational dropout (i.e. Hetero+Var.(I), (II)) leads to the top 2 performance on both datasets on interior and reduces errors on exterior to the level comparable or better than the baseline.

The performance difference of heteroscedastic network between interior and exterior region roots from the loss function. The term $\mathcal{M}_\theta(\mathcal{D})$ imposes a larger penalty on regions with smaller intrinsic uncertainty. The network thus allocates more resources towards the lower noise regions where the statistical mapping from the LR to HR space is less ambiguous. The dramatic reduction on the exterior error from variational dropout indicates its regularisation effect against such overfitting, and also improves the robustness of prediction on the interior.

**Tractography with MAP-MRI SR:** Reconstruction accuracy does not reflect real world utility. We thus further assessed SR quality with a tractography experiment on the Prisma dataset, which contains two DWIs of the same subject from a Siemens 3T scanner, with 1.35 mm and 2.5 mm resolution. An ensemble of 8 hetero+var.(I) CNNs super-resolves the MAP-MRI coefficients [9] derived from the LR DWIs (2.5 mm), then the HR MAP volume is used to predict the HR DWIs (1.25 mm). The final prediction is computed as the average estimate weighted by the inverse covariance as in RF-IQT.

Fig. 4 shows streamline maps of the probabilistic tractography [13] for the original LR/HR data and various upsampled images, and focuses on examples that highlight the benefits of reduced SR reconstruction errors. In the top row, tractography on the LR data produces a false-positive tract under the corpus callosum (yellow arrow in the 1st row), which tractography at HR avoids. Reconstructured HR images from IQT-RF and CNN avoid the false positive better than linear intepolation. Note that we do not expect to reproduce the HR trac-

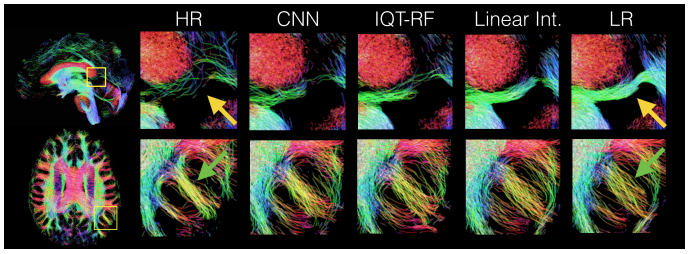


**Fig. 4:** Tractography on Prisma dataset for different methods. From left to right: (i) HR acquisition, (ii) CNN prediction; (iii) RF; (iv) Linear interpolation; (v) LR acquisition.

(a) Uncertainty propagation (b) Tumour

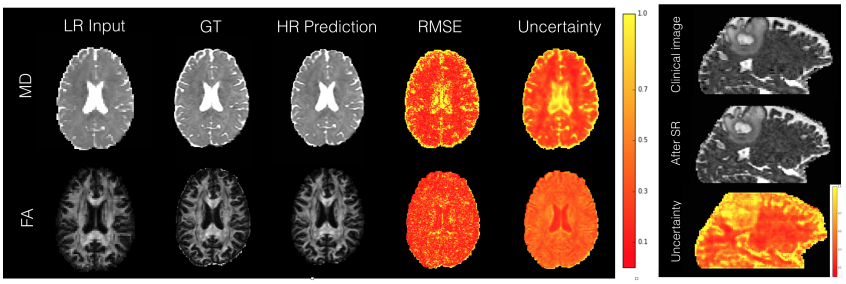


**Fig. 5:** (a) Comparison between RMSE and uncertainty maps for FA and MD computed on a HCP subject. LR input, ground truth and HR prediction are also shown. (b) DTI SR on a brain tumour patient. From top to bottom: (i) MD computed from the original DTI; (ii) the estimated HR version; (iii) uncertainty.

tography map exactly, as the HR and LR images are not perfectly aligned. The bottom row shows shaper recovery of small gyral white matter pathways (green arrow) at HR than LR resulting from reduced partial volume effect. CNN reconstruction produces a sharper pathway than RF-IQT and linear interpolation.

**Visualisation of predictive uncertainty:** We measure the expectation and variance of *mean diffusivity* (MD) and *fractional anisotropy* (FA) with respect to the predictive distribution $q^*_\phi(\mathbf{y}|\mathbf{x})$ of Hetero+Var.(I) by MC sampling. Comparative results are shown in Fig. 5(a), where we drew 200 samples of HR DTIs from the predictive distribution. The uncertainty map is highly correlated with the error maps. In particular, the MD uncertainty map captures subtle variations within the white matter and central CSF, which demonstrates its potential utility of the uncertainty map as a surrogate measure of accuracy. Fig. 5(b) shows the same model trained on a healthy HCP cohort applied to the DTI of a brain tumour patient. The raw data (DWI) with $b = 700$ s/mm$^2$ is processed as before with input voxel size $2^3$ mm$^3$. The ground truth is unavailable but the estimated image sharpens the input without introducing noticeable artifacts. Uncertainty is high on the tumour, not represented in the training data, again illustrating its potential to flag plausible low accuracy areas.

## 4 Discussion

We present a super-resolution algorithm based on 3D subpixel-CNNs with state-of-the-art accuracy and efficiency on diffusion MRI datasets. An application to the MAP-MRI coefficients indicates benefits to tractography. We also show that assimilation of *intrinsic* and *parameter* uncertainty in the model leads to best predictive performance. The uncertainty map highly correlates with reconstruction errors and is able to highlight pathologies. Understanding the behaviours of these uncertainty measures in unfamiliar test environments (e.g. pathologies) and their relations to predictive performance is an important future work for designing a more generalisable method. The presented ideas extend to many other quality enhancement problems in medical image analysis and beyond.

**Acknowledgements.** This work was supported by Microsoft scholarship. Data were provided in part by the HCP, WU-Minn Consortium (PIs: David Van Essen and Kamil Ugurbil) funded by NIH and Wash. U. The tumour data were acquired as part of a study lead by Alberto Bizzi, MD at his hospital in Milan, Italy.

## References

1. Alexander, D.C., et al.: Image quality transfer and applications in diffusion MRI. NeuroImage **152** (2017) 283–298
2. Oktay, O., et al.: Multi-input cardiac image super-resolution using convolutional neural networks. In: MICCAI, Springer (2016)
3. Bahrami, K., Shi, F., Rekik, I., Shen, D.: Convolutional neural network for reconstruction of 7T-like images from 3T MRI using appearance and anatomical features. In: MICCAI DLDLM workshop, Springer (2016) 39–47
4. Tanno, R., Ghosh, A., Grussu, F., Kaden, E., Criminisi, A., Alexander, D.C.: Bayesian image quality transfer. In: MICCAI, Springer (2016) 265–273
5. Shi, W., et al.: Real-time single image and video super-resolution using an efficient sub-pixel convolutional neural network. In: CVPR. (2016) 1874–1883
6. Nix, D.A., Weigend, A.S.: Estimating the mean and variance of the target probability distribution. In: IEEE WCCI. Volume 1., IEEE (1994) 55–60
7. Kingma, D.P., Salimans, T., Welling, M.: Variational dropout and the local reparameterization trick. In: NIPS. (2015) 2575–2583
8. Sotiropoulos, S.N., et al.: Advances in diffusion MRI acquisition and processing in the human connectome project. Neuroimage **80** (2013) 125–143
9. Özarslan, et al.: Mean apparent propagator (MAP) MRI: a novel diffusion imaging method for mapping tissue microstructure. NeuroImage **78** (2013) 16–32
10. Dong, C., Loy, C.C., He, K., Tang, X.: Image super-resolution using deep convolutional networks. IEEE PAMI **38**(2) (2016) 295–307
11. Johnson, J., Alahi, A., Fei-Fei, L.: Perceptual losses for real-time style transfer and super-resolution. In: ECCV, Springer (2016) 694–711
12. Kingma, D.P., Ba, J.: Adam: A method for stochastic optimization. CoRR **abs/1412.6980** (2014)
13. Tournier, J., Calamante, F., Connelly, A.: Improved probabilistic streamlines tractography by 2nd order integration over fibre orientation distributions. In: ISMRM. (2010) 1670