


Bayesian Image Quality Transfer

Ryutaro Tanno^{1,3}, Aurobrata Ghosh¹, Francesco Grussu², Enrico Kaden¹, Antonio Criminisi³, and Daniel C. Alexander¹

¹ Centre for Medical Image Computing, University College London, London, UK
ryutaro.tanno.15@ucl.ac.uk

² Institute of Neurology, University College London, London, UK

³ Machine Intelligence and Perception Group,
Microsoft Research Cambridge, Cambridge, UK

Abstract. Image quality transfer (IQT) aims to enhance clinical images of relatively low quality by learning and propagating high-quality structural information from expensive or rare data sets. However, the original framework gives no indication of confidence in its output, which is a significant barrier to adoption in clinical practice and downstream processing. In this article, we present a general Bayesian extension of IQT which enables efficient and accurate quantification of uncertainty, providing users with an essential prediction of the accuracy of enhanced images. We demonstrate the efficacy of the uncertainty quantification through super-resolution of diffusion tensor images of healthy and pathological brains. In addition, the new method displays improved performance over the original IQT and standard interpolation techniques in both reconstruction accuracy and robustness to anomalies in input images.

AQ1
AQ2

1 Introduction

A diverse range of technological and economical factors constrain the quality of magnetic resonance (MR) images. Whilst there exist bespoke scanners or imaging protocols with the capacity to generate ultra high-quality data, their prohibitive cost and lengthy acquisition time render the technology impractical in clinical applications. On the other hand, the poor quality of clinical data often limits the accuracy of subsequent analysis. For example, low spatial resolution of diffusion weighted images (DWI) gives rise to partial volume effects, introducing a bias in diffusion tensor (DT) measurements [1] that are widely used to study white matter in terms of anatomy, neurological diseases and surgical planning.

Super-resolution (SR) reconstruction potentially addresses this challenge by post-processing to increase the spatial resolution of a given low-resolution (LR) image. One popular approach is the single-image SR method, which attempts to recover a high-resolution (HR) image from a single LR image. Numerous machine-learning based methods have been proposed. For instance, [2, 3] use example patches from HR images to super-resolve scalar MR and DW images respectively, with an explicitly defined generative model relating a HR patch to a LR patch and carefully crafted regularisation. Another generative approach is the

sparse-representation methods [4, 5], which construct a coupled library of HR and LR images from training data and solve the SR problem through projection onto it. Image quality transfer (IQT) [6] is a general quality-enhancement framework based on patch regression, which shows great promise in SR of DT images and requires no special acquisition, so is applicable to large varieties of existing data.

A key limitation of above methods is the lack of a mechanism to communicate confidence in the predicted HR image. High quality training data typically come from healthy volunteers. Thus, performance in the presence of pathology or other effects not observed in the training data is questionable. We expect methods to have high confidence in regions where the method has seen lots of similar examples during training, and lower confidence on previously unseen structures. However, current methods implicitly have equal confidence in all areas. Such an uncertainty characterisation is particularly important in medical applications where ultimately images can inform life-and-death decisions. It is also beneficial to downstream image processing algorithms, such as registration or tractography.

In this paper, we extend the IQT framework to predict and map uncertainty in its output. We incorporate Bayesian inference into the framework and name the new method Bayesian IQT (BIQT). Although many SR methods [2–5] can be cast as *maximum a posteriori* (MAP) optimisation problems, the dimensionality or complexity of the posterior distribution make the computation of uncertainty very expensive. In contrast, the random forest implementation of the original IQT is amenable to uncertainty estimation thanks to the simple linear model at each leaf node, but the current approach computes *maximum likelihood* (ML) solution. BIQT replaces this ML based inference with Bayesian inference (rather than just MAP) and this allows the uncertainty estimate to reflect unfamiliarity of input data (see Fig. 1(a)). We demonstrate improved performance through SR of DT images on Human Connectome Project (HCP) dataset [7], which has sufficient size and resolution to provide training data and a testbed to gauge the baseline performance. We then use clinical data sets from multiple sclerosis (MS) and tumour studies to show the efficacy of the uncertainty estimation in the presence of focal tissue damage, not represented in the HCP training data.

2 Methods

Here we first review the original IQT framework based on a regression forest. We then introduce our Bayesian extension, BIQT, highlighting the proposed efficient hyperparameter optimisation method and the robust uncertainty measure.

Background. IQT splits a LR image into small patches and performs quality enhancement on them independently. This patch-wise reconstruction is formulated as a regression problem of learning a mapping from each patch \mathbf{x} of N_l voxels in the LR image to a corresponding patch $\mathbf{y}(\mathbf{x})$ of N_h voxels in the HR image. Input and output voxels are vector-valued containing p_l and p_h values, and thus the mapping is $\mathbf{x} \in \mathbf{R}^{N_l p_l} \rightarrow \mathbf{y}(\mathbf{x}) \in \mathbf{R}^{N_h p_h}$. Training data comes from high quality data sets, which are artificially downsampled to provide matched pairs of LR and HR patches. For application, each patch of a LR image is passed

through the learned mapping to obtain a HR patch and those patches combine to estimate a HR image. To solve the regression problem, IQT employs a variant of random forests [8]. The method proceeds in two stages: training and prediction.

During training, we grow a number of trees on different sets of training data. Each tree implements a piecewise linear regression; it partitions the input space $\mathbf{R}^{N_{lpt}}$ and performs regressions in respective subsets. Learning the structure of a tree on dataset $\mathcal{D} = \{\mathbf{x}_i, \mathbf{y}_i\}_{i=1}^{|\mathcal{D}|}$ aims to find an ‘optimal’ sequence of the following form of binary partitioning. At the initial node (root), \mathcal{D} is split into two sets \mathcal{D}_R and \mathcal{D}_L by thresholding one of J scalar functions of \mathbf{x} , or *features*, f_1, \dots, f_J . The optimal pair of a feature f_m and a threshold τ with the most effective splitting is selected by maximising the *information gain* [9], $IG(f_m, \tau, \mathcal{D}) \triangleq |\mathcal{D}| \cdot H(\mathcal{D}) - |\mathcal{D}_R| \cdot H(\mathcal{D}_R) - |\mathcal{D}_L| \cdot H(\mathcal{D}_L)$ where $|\mathcal{D}|$ denotes the size of set \mathcal{D} and $H(\mathcal{D})$ is the average differential entropy of the predictive distribution $P(\mathbf{y}|\mathbf{x}, \mathcal{D}, \mathcal{H})$ given by

$$H(\mathcal{D}) \triangleq -\frac{1}{|\mathcal{D}|} \sum_{\mathbf{x} \in \mathcal{D}} \int P(\mathbf{y}|\mathbf{x}, \mathcal{D}, \mathcal{H}) \cdot \log P(\mathbf{y}|\mathbf{x}, \mathcal{D}, \mathcal{H}) \, d\mathbf{y}. \quad (1)$$

Maximising the information gain helps selecting the splitting with highest confidence in predictive distributions. This optimization problem is solved by performing golden search on the threshold for all features. The hypothesis space \mathcal{H} specifies the class of statistical models and governs the form of predictive distribution. In particular, IQT fits the ML estimation of a linear model with a Gaussian noise. To control over-fitting, a validation set \mathcal{D}^V with similar size to \mathcal{D} is used and the root node is only split if the residual error is reduced. This process is repeated in all new nodes until no more splits pass the validation test.

At the time of prediction, every LR patch \mathbf{x} is routed to one of the leaf nodes (nodes with no children) in each tree through a series of binary splitting learned during training, and the corresponding HR patch is estimated by the mode of the predictive distribution. The forest output is computed as the average of predictions from all trees weighted by the inverted variance of predictive distributions.

Bayesian Image Quality Transfer. Our method, BIQT follows the IQT framework described above and performs a patch-wise reconstruction using a regression forest. The key novelty lies in our Bayesian choice of \mathcal{H} (Eq. (1)). For a given training set at a node, $\mathcal{D} = \{\mathbf{x}_i, \mathbf{y}_i\}_{i=1}^N$, BIQT fits a Bayesian linear model $\mathbf{y} = \mathbf{W}\mathbf{x} + \boldsymbol{\eta}$ where the additive noise $\boldsymbol{\eta}$ and the linear transform $\mathbf{W} \in \mathbf{R}^{N_{hp} \times N_{lpt}}$ follow isotropic Gaussian distributions $P(\boldsymbol{\eta}|\beta) = \mathcal{N}(\boldsymbol{\eta}|\mathbf{0}, \beta^{-1}\mathbf{I})$ and $P(\mathbf{W}_l|\alpha) = \mathcal{N}(\mathbf{W}_l|\mathbf{0}, \alpha^{-1}\mathbf{I})$, with \mathbf{W}_l denoting the row-wise vectorised version of \mathbf{W} . The hyperparameters α and β are positive scalars, and \mathbf{I} denotes an identity matrix. Assuming for now that α, β are known, the predictive distribution is computed by marginalising out the model parameters \mathbf{W} as

$$P(\mathbf{y}|\mathbf{x}, \mathcal{D}, \mathcal{H}) = P(\mathbf{y}|\mathbf{x}, \mathcal{D}, \alpha, \beta) = \mathcal{N}(\mathbf{y}|\mathbf{W}_{\text{Pred}}\mathbf{x}, \sigma_{\text{Pred}}^2(\mathbf{x}) \cdot \mathbf{I}) \quad (2)$$

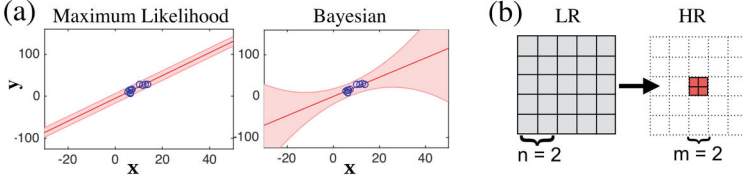


Fig. 1. (a) 1D illustration (i.e. both $\mathbf{x}, \mathbf{y} \in \mathbf{R}$) of ML and Bayesian linear models fitted to the data (blue circles). The red line and shaded areas show the mode and variance (uncertainty) of $P(\mathbf{y}|\mathbf{x}, \mathcal{D}, \mathcal{H})$ at respective \mathbf{x} values. Bayesian method assigns high uncertainty to an input distant from the training data whilst the ML's uncertainty is fixed. (b) 2D illustration of the input (grey) and output (red) patches.

where the i^{th} columns of matrices \mathbf{X} and \mathbf{Y} are given by \mathbf{x}_i and \mathbf{y}_i , the mean linear map $\mathbf{W}_{\text{Pred}} = \mathbf{Y}\mathbf{X}^T(\mathbf{X}\mathbf{X}^T + \frac{\alpha}{\beta}\mathbf{I})^{-1}$ and the variance $\sigma_{\text{Pred}}^2(\mathbf{x}) = \mathbf{x}^T \mathbf{A}^{-1} \mathbf{x} + \beta^{-1}$ with $\mathbf{A} = \alpha \mathbf{I} + \beta \mathbf{X}\mathbf{X}^T$. The mean differential entropy in Eq. (1) can be computed as $H(\mathcal{D}) = N_h p_h |\mathcal{D}|^{-1} \sum_{\mathbf{x} \in \mathcal{D}} \log(\mathbf{x}^T \mathbf{A}^{-1} \mathbf{x} + \beta^{-1})$ (up to constants).

The predictive variance $\sigma_{\text{Pred}}^2(\mathbf{x})$ provides an informative measure of uncertainty over the enhanced patch $\mathbf{y}(\mathbf{x})$ by combining two quantities: the degree of variation in the training data, β^{-1} and the degree of ‘familiarity’, $\mathbf{x}^T \mathbf{A}^{-1} \mathbf{x}$ which measures how different the input patch \mathbf{x} is from the observed data. For example, if \mathbf{x} contains previously unseen features such as pathology, the familiarity term becomes large, indicating high uncertainty. The equivalent measure for the original IQT, however, solely consists of the term, β^{-1} determined from the training data, and yields a fixed uncertainty estimate for any new input \mathbf{x} (see Fig. 1(a)). Once a full BIQT forest \mathcal{F} is grown, we perform reconstruction in the same way as before. All leaf nodes are endowed with the predictive distributions of the form in Eq. (2), and BIQT quantifies the uncertainty over the HR output, $\mathbf{y}(\mathbf{x})$ as the predictive variance at leaf nodes (at which \mathbf{x} arrives after traversing respective trees) $\langle \sigma_{\text{Pred}}^2(\mathbf{x}) \rangle_{\mathcal{F}}$ averaged over trees in the forest \mathcal{F} .

A priori the hyper-parameters α and β are unknown, so we optimise them by maximising the marginal likelihood $P(\mathcal{D}|\alpha, \beta)$. As \mathbf{W}_{Pred} is in fact a solution of L2 regularisation problem with smoothing α/β , this optimisation procedure can be viewed as a data-driven determination of regularisation level. Although a closed form for $P(\mathcal{D}|\alpha, \beta)$ exists, exhaustive search is impractical as we have to solve this problem for every binary splitting (characterised by a feature and a threshold) at all internal nodes of the tree. We thus derive and use the multi-output generalisation of the Gull-Mackay fixed-point iteration algorithm [10]

$$\beta_{\text{new}} = \frac{1 - \beta_{\text{old}} \cdot |\mathcal{D}|^{-1} \text{trace}(\mathbf{A}(\alpha_{\text{old}}, \beta_{\text{old}})^{-1} \mathbf{X}\mathbf{X}^T)}{\frac{1}{|\mathcal{D}| N_h p_h} \sum_{j=1}^{N_h p_h} \sum_{i=1}^{|\mathcal{D}|} [y_{ji} - \boldsymbol{\mu}_j(\alpha_{\text{old}}, \beta_{\text{old}})^T \mathbf{x}_i]^2} \quad (3)$$

$$\alpha_{\text{new}} = \frac{N_l p_l - \alpha_{\text{old}} \cdot \text{trace}(\mathbf{A}(\alpha_{\text{old}}, \beta_{\text{old}})^{-1})}{\frac{1}{N_h p_h} \sum_{j=1}^{N_h p_h} \boldsymbol{\mu}_j(\alpha_{\text{old}}, \beta_{\text{old}})^T \boldsymbol{\mu}_j(\alpha_{\text{old}}, \beta_{\text{old}})} \quad (4)$$

where $\mu_j(\alpha, \beta) = \beta \cdot \mathbf{A}(\alpha, \beta)^{-1} \sum_{i=1}^D y_{ji} \mathbf{x}_i$. Whilst the standard MATLAB optimisation solver (e.g. `fminunc`) requires at least 50 times more computational time per node optimisation than for IQT, this iterative method is only average 2.5 times more expensive, making the Bayesian extension viable. We use this over Expectation Maximisation algorithm for its twice-as-fast convergence rate.

3 Experiments and Results

Here we demonstrate and evaluate BIQT through the SR of DTI. First we describe the formulation of the application. Second, we compare the baseline performance on the HCP data to the original IQT. Lastly, we demonstrate on clinical images of diseased brains that our uncertainty measure highlights pathologies.

Super-Resolution of DTIs. Given a LR image, BIQT enhances its resolution patch by patch. Each case takes as input a $N_l = (2n + 1)^3$ cubic patch of voxels each containing $p_l = 6$ DT elements, and super-resolves its central voxel by a factor of m (so the output is a $N_h = m^3$ patch with each new voxel also containing $p_h = 6$ DT components). For all the experiments, we use $n = 2$ and $m = 2$ (Fig. 1(b)) and so the map is $\mathbf{R}^{750=5^3 \times 6} \rightarrow \mathbf{R}^{48=2^3 \times 6}$. The features $\{f_i\}$ consist of mean eigenvalues, principal orientation, orientation dispersion averaged over central subpatches of different widths = 1, 3, 5 within the LR patch.

Training data is generated from 8 randomly selected HCP subjects and used for all subsequent experiments. We use a subsample of each dataset, which consists of 90 DWIs of voxel size 1.25^3 mm^3 with $b = 1000 \text{ s/mm}^2$. We create training pairs by downsampling each DWI by a factor of m and then fitting the DT to the downsampled and original DWI. A coupled library of LR and HR patches is then constructed by associating each patch in the downsampled DTI with the corresponding patch in the ground truth DTI. Training of each tree is performed on a different data set obtained by randomly sampling $\approx 10^5$ pairs from this library, and it takes under 2 h for the largest data sets in Fig. 2.

Testing on HCP dataset. We test BIQT on another set of 8 subjects from the HCP cohort. To evaluate reconstruction quality, three metrics are used: the root-mean-squared-error of the six independent DT elements (DT RMSE); the Peak Signal-to-Noise Ratio (PSNR); and the mean Structural Similarity (MSSIM) index [11]. We super-resolve each DTI after downsampling by a factor of 2 as before, and these quality measures are then computed between the reconstructed HR image and the ground-truth. BIQT displays highly statistically significant ($p < 10^{-8}$) improvements (see Fig. 2) on all three metrics over IQT, linear regression methods and a range of interpolation techniques. In addition, trees obtained with BIQT are generally deeper than those of the original IQT.

Standard linear regression performs as well as the Bayesian regression due to the large training data size. However, with BIQT, as you descend each tree, the number of training data points at each node gets smaller, increasing the degree

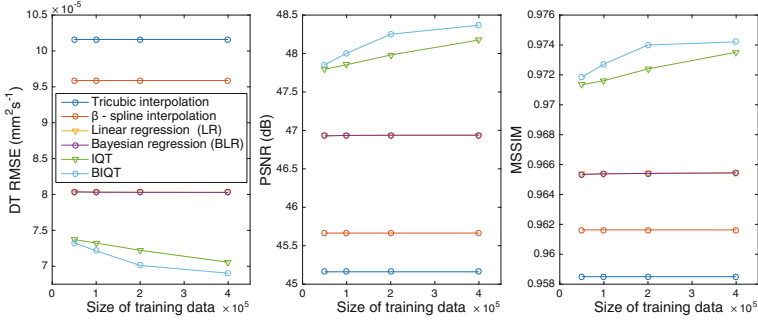


Fig. 2. Three reconstruction metrics of various SR methods as a function of training data size; RMSE (left), PSNR (middle) and MSSIM (right). The performance of LR (yellow) and BLR (purple) coincide. The results for linear and nearest-neighbour interpolation are omitted for their poor performance.

of uncertainty in model fitting, and so the data-driven regularisation performed in each node-wise Bayesian regression becomes more effective, leading to better reconstruction quality. This is also manifested in the deeper structure of BIQT trees, indicating more successful validation tests and thus greater generalisability. Moreover, the feedforward architectures of trees and parallelisability of patch-wise SR means highly efficient reconstruction (a few minutes for a full volume).

Figure 3 shows reconstruction accuracies and uncertainty maps for BIQT and IQT. The uncertainty map of BIQT is more consistent with its reconstruction accuracy when compared to the original IQT. Higher resemblance is also observed between the distribution of accuracy (RMSE) and uncertainty (variance). The BIQT uncertainty map also highlights subtle variations in the reconstruction-quality within the white matter, whereas the IQT map contains flatter contrasts with discrete uncertainties that vary greatly in the same region (see histograms in bottom row). This improvement reflects the positive effect of the data-driven regularisation and better generalisability of BIQT and can be observed particularly in the splenium and genu of the Corpus Callosum, where despite good reconstruction accuracy, IQT assigns higher uncertainty than in the rest of the white matter and BIQT indicates a lower and more consistent uncertainty. Thus, the BIQT uncertainty map displays higher correspondence with accuracy and allows for a more informative assessment of reconstruction quality. Note that while the uncertainty measure for IQT is governed purely by the training data, for BIQT the uncertainty also incorporates the familiarity of the test data.

Testing on MS and tumour data. We further validate our method on images with previously unseen abnormalities; we use trees trained on healthy subjects from HCP dataset to super-resolve DTIs of MS and brain tumour patients (10 each). We process the raw data (DWI) as before, and only use $b = 1200$ s/mm² measurements for the MS dataset and $b = 700$ s/mm² for the tumour dataset.

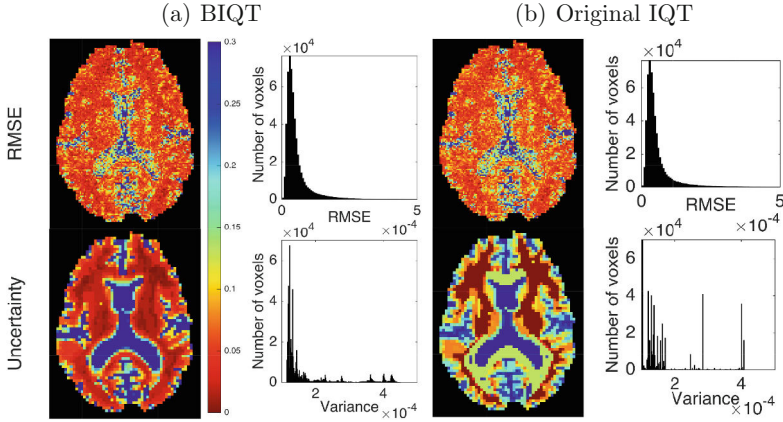


Fig. 3. Reconstruction accuracy and uncertainty maps. (top row) The voxel-wise RMSE as a normalised colour-map and its distribution; (bottom row) Uncertainty map (variance) over the super-resolved voxels and its distribution for (a) BIQT and (b) IQT. Trees were trained on $\approx 4 \times 10^5$ patch pairs.

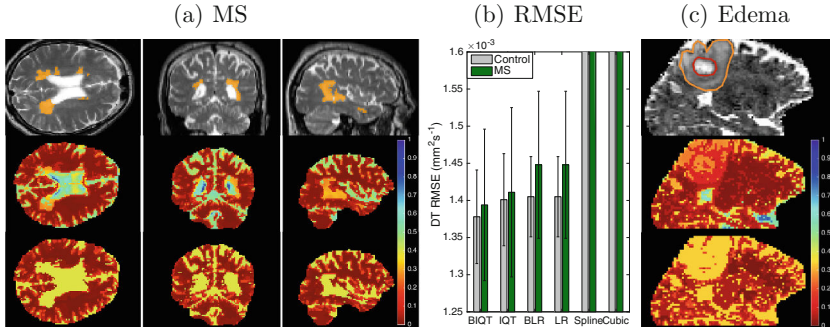


Fig. 4. (a), (c) Normalised uncertainty map (variance is shown i.e. the smaller the more certain) for BIQT (middle row) and IQT (bottom row) along with the T2-weighted slices (top row) for MS (with focal lesions in orange) and edema (contours highlighted), respectively. (b). The RMSE for MS and control subjects (averaged over 10 subjects in each case).

The voxel size for both datasets is 2^3 mm^3 . The MS dataset also contains lesion masks manually outlined by a neurologist. Figure 4(a), (c) middle row shows that the uncertainty map of BIQT precisely identifies previously unseen features (pathologies in this case) by assigning lower confidence than for the remaining healthy white matter. Moreover, in accordance with the reconstruction accuracy, the prediction is more confident in pathological regions than in the cerebrospinal fluid (CSF). This is expected since the CSF is essentially free water with low SNR and is also affected by cardiac pulsations, whereas the pathological regions are contained within the white matter and produce better SNR. Each BIQT tree

appropriately sends pathological patches into the ‘white-matter’ subspace and its abnormality is detected there by the ‘familiarity’ term, leading to a lower confidence with respect to the healthy white matter. By contrast, IQT sends pathological patches into the CSF subspace and assigns the fixed corresponding uncertainty which is higher than what it should be. In essence, BIQT enables an uncertainty measure which highly correlates with the pathologies in a much more plausible way, and this is achieved by its more effective partitioning of the input space and uncertainty estimation conferred by Bayesian inference. Moreover, Fig. 4(b) shows the superior generalisability of BIQT even in reconstruction accuracy (here SR is performed on downsampled clinical DTIs); the RMSE of BIQT for MS patients is even smaller than that of IQT for healthy subjects.

4 Conclusion

We presented a computationally viable Bayesian extension of Image Quality Transfer (IQT). The application in super resolution of DTI demonstrated that the method not only achieves better reconstruction accuracy even in the presence of pathology (Fig. 3(b)) than the original IQT and standard interpolation techniques, but also provides an uncertainty measure which is highly correlated with the reconstruction quality. Furthermore, the uncertainty map is shown to highlight focal pathologies not observed in the training data. BIQT also performs a computationally efficient reconstruction while preserving the generality of IQT with large potential to be extended to higher-order models beyond DTI and applied to a wider range of modalities and problems such as parameter mapping and modality transfer [6]. We believe that these results are sufficiently compelling to motivate larger-scale experiments for clinical validation in the future.

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; 1U54MH091657) 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. The MS data were acquired as part of a study at UCL Institute of Neurology, funded by the MS Society UK and the UCL Hospitals Biomedical Research Centre (PIs: David Miller and Declan Chard).

References

1. Berlot, R., Metzler-Baddeley, C., Jones, D.K., O’Sullivan, M.J., et al.: CSF contamination contributes to apparent microstructural alterations in mild cognitive impairment. *Neuroimage* **92**, 27–35 (2014)
2. Rousseau, F.: Brain hallucination. In: Forsyth, D., Torr, P., Zisserman, A. (eds.) *ECCV 2008*. LNCS, vol. 5302, pp. 497–508. Springer, Heidelberg (2008). doi:[10.1007/978-3-540-88682-2_38](https://doi.org/10.1007/978-3-540-88682-2_38)
3. Coupé, P.: Collaborative patch-based super-resolution for diffusion-weighted images. *Neuroimage* **83**, 245–261 (2013)

4. Rueda, A., Malpica, N., Romero, E.: Single-image super-resolution of brain MR images using overcomplete dictionaries. *Med. Image Anal.* **17**(1), 113–132 (2013)
5. Wang, Y.H., Qiao, J., Li, J.B., Fu, P., Chu, S.C., Roddick, J.F.: Sparse representation-based MRI super-resolution reconstruction. *Measurement* **47**, 946–953 (2014)
6. Alexander, D.C., Zikic, D., Zhang, J., Zhang, H., Criminisi, A.: Image quality transfer via random forest regression: applications in diffusion MRI. In: Golland, P., Hata, N., Barillot, C., Hornegger, J., Howe, R. (eds.) *MICCAI 2014. LNCS*, vol. 8675, pp. 225–232. Springer, Heidelberg (2014). doi:[10.1007/978-3-319-10443-0_29](https://doi.org/10.1007/978-3-319-10443-0_29)
7. Sotiropoulos, S.N., et al.: Advances in diffusion MRI acquisition and processing in the human connectome project. *Neuroimage* **80**, 125–143 (2013)
8. Breiman, L.: Random forests. *Mach. Learn.* **45**(1), 5–32 (2001)
9. Criminisi, A., Shotton, J.: *Decision Forests for Computer Vision and Medical Image Analysis*. Springer, London (2013)
10. MacKay, D.J.: Bayesian interpolation. *Neural Comput.* **4**(3), 415–447 (1992)
11. Wang, Z., Bovik, A.C., Sheikh, H.R., Simoncelli, E.P.: Image quality assessment: from error visibility to structural similarity. *IEEE Trans. Image Proc.* **13**(4), 600–612 (2004)