

Image quality transfer and applications in diffusion MRI

Daniel C. Alexander¹ Darko Zikic² Aurobrata Ghosh¹
Ryutaro Tanno¹ Viktor Wottschel³ Jiaying Zhang¹
Enrico Kaden¹ Tim B. Dyrby^{4,5}
Stamatios N. Sotiropoulos⁶ Hui Zhang¹
Antonio Criminisi²

¹Centre for Medical Image Computing and Dept. Computer Science,
UCL, Gower Street, London, WC1E 6BT.

²Microsoft Research Cambridge, Cambridge, UK.

³Institute of Neurology, UCL, Queen Square, London.

⁴Danish Research Centre for Magnetic Resonance,
Centre for Functional and Diagnostic Imaging and Research,
Copenhagen University Hospital Hvidovre, Hvidovre, Denmark.

⁵Dept. Applied Maths and Computer Science,
Technical University of Denmark, Lyngby, Denmark.

⁶FMRIB Centre, University of Oxford, John Radcliffe Hospital,
Headington, UK.

February 19, 2017

Abstract

This paper introduces a new computational imaging technique called image quality transfer (IQT). IQT uses machine learning to transfer the rich information available from one-off experimental medical imaging devices to the abundant but lower-quality data from routine acquisitions. The procedure uses matched pairs to learn mappings from low-quality to corresponding high-quality images. Once learned, these mappings then augment unseen low quality images, for example by enhancing image resolution or information content. Here, we demonstrate IQT using a simple patch-regression implementation and the uniquely rich diffusion MRI data set from the human connectome project (HCP). **Results highlight potential benefits** of IQT in both brain connectivity mapping and microstructure imaging. In brain connectivity mapping, IQT reveals, from standard data sets, thin connection pathways that tractography normally requires specialised data to reconstruct. **In microstructure imaging, IQT shows potential in estimating,** from standard “single-shell” data (one non-zero b -value), maps of microstructural parameters that normally require specialised multi-shell data. Further experiments show strong generalisability, highlighting IQT’s benefits even when the training set does not directly represent the application domain. The concept extends naturally to many other imaging modalities and reconstruction problems. **R2.1** **R2.1**

Introduction

Bespoke MRI scanners and imaging protocols can produce very high quality data uniquely informative on anatomy and physiology. For example, the Human Connectome Project (HCP) developed specialised human MRI scanners fitted with 100mTm^{-1} and 300mTm^{-1} gradient coils [1, 2], which are up to an order of magnitude more powerful than standard clinical scanners. In combination with custom imaging and image reconstruction innovations [3] and a lengthy acquisition protocol, these devices provide unprecedented image resolution, signal levels, and coverage of the measurement space revealing ultra-fine anatomical detail from live subjects. Such experimental devices showcase the potential for future clinical imaging systems [4], but provide little immediate benefit to current clinical practice where hardware is more modest and patient-imaging time is very limited. Moreover, translation is slow, because market forces demand strong evidence that mass production has major benefit. Even when available, the latest technology is often prohibitively expensive outside elite research centres, so the state of the art remains inaccessible to most practitioners.

Image quality transfer (IQT) aims to bridge this gap between rare experimental systems and accessible commercial systems. The technique learns mappings from low quality (e.g. clinical) to high quality (e.g. experimental) images exploiting the similarity of images across subjects, regions, modalities, and scales: image macro- and meso-structure is highly predictive of sub-voxel content. The mapping may then operate directly on low-quality images to estimate the corresponding high-quality images, or serve as a prior in an otherwise ill-posed image-reconstruction routine.

Such a procedure has the potential to enhance or enable a wide range of desirable

imaging or image analysis operations. These include long-standing challenges in medical imaging, such as enhancement of spatial resolution [5, 6, 7, 8, 9] and harmonisation of multi-centre studies [10] or longitudinal studies that straddle scanner upgrades, as well as more recent developments, such as image reconstruction from sparse acquisition (compressed sensing or MR fingerprinting) [11, 12], or predicting one contrast or modality from another (“image synthesis” or “modality propagation”) [13, 14, 15, 16].

We construct two simple and practical implementations of IQT via patch-regression using i) a global linear transformation (IQT-GL), and ii) a non-linear random-forest (IQT-RF). To demonstrate the concept and benefits of IQT, we focus on using it to enhance spatial resolution and the downstream benefits to brain connectivity mapping. We also investigate the potential for IQT to enrich the information content of images by estimating maps of microstructural indices that normally rely on “multi-shell” acquisition but from “single-shell” input.

Image resolution is a major limiting factor in tractography. IQT can reconstruct high resolution images from low resolution input by learning a mapping from low resolution to high resolution images. We demonstrate here that this operation improves dramatically on interpolation, the standard technique that has well-documented limitations [17], and enables tractography to identify white matter pathways in low-resolution images that otherwise require specialised high-resolution data. Diffusion MRI super-resolution techniques, e.g. [6, 7, 8, 9], potentially offer similar benefits. However, most rely on a specialised acquisition with partially overlapping anisotropic voxel grids, cf. [5], and/or involve complex processing pipelines [6], so are not widely used. In contrast, IQT operates naturally on any existing data set and requires no special acquisition. In fact, the IQT operation is complementary to the reconstruction procedures in [7, 8, 9].

Additional experiments aim to recover parameter maps provided by neurite orientation dispersion and density imaging (NODDI) [18] and microscopic diffusion anisotropy mapping based on the spherical mean technique (SMT) [19, 20] from single-shell data sets, i.e. with only one non-zero b -value. NODDI maps indices of neurite (axons and dendrites) density and their geometric configuration (orientation dispersion); SMT [19] maps the per-axon microscopic diffusion tensor independent of intra-voxel orientation distribution, as well as orientation dispersion. Both exemplify the microstructure-imaging paradigm [21], but require non-standard multi-shell acquisitions with multiple non-zero b -values and fail given single-shell data sets, which are routinely acquired for diffusion tensor imaging (DTI) [22]. IQT learns mappings from single-shell (DTI) to multi-shell (NODDI/SMT) parameter maps. The ability to estimate such parameters from single-shell input data potentially enables reanalysis of a wide range of historical brain-imaging studies to reveal more specific information on group differences or disease mechanisms. The closest previous work on this topic [23] learns a prior from examples to stabilise estimates of advanced diffusion MRI indices, specifically NODDI parameters and diffusion kurtosis [24], from sparse data sets. However, the data sets in [23] are still multi-shell so the parameter-estimation problem remains well posed.

Methods

This section first specifies the various data sets and diffusion MRI models the experiments in the next section use to demonstrate IQT. We then describe the general implementation of IQT through patch regression and detail the specific versions designed for resolution enhancement and parameter mapping.

Data sets

We make use of four data sets to test and evaluate our implementations of IQT.

HCP data set

The main data set for training and testing is the diffusion MRI data provided by the WU-Minn HCP, release Q3 [25]. The data from each subject includes 288 diffusion weighted images (DWIs) (acquired twice with reversed phase encoding direction) with 1.25 mm isotropic voxels; 18 have nominal $b = 0$ and the three high-angular-resolution-diffusion-imaging (HARDI) shells of 90 directions have nominal b -values of 1000, 2000, and 3000 s mm⁻²; the precise values vary spatially, see [3], which fully details the acquisition. The data are corrected for distortions (susceptibility-induced, eddy currents and subject motion), see [26].

The HCP cohort contains broadly healthy, young-adult (22-36 years) subjects, who vary in race, gender, and handedness all of which affect brain structure. We identify various subgroups of the HCP cohort to evaluate generalisability. Subgroup 1, which we refer to as HCP1, is as heterogeneous as possible containing a mixture of “White - Hispanic/Latino”, “White - Not Hispanic”, and “Black or African Am.”, unrelated, male and female, and left and right-handed subjects in their 20s and 30s. Subgroups 2 (HCP2) and 3 (HCP3) are as homogeneous as possible while being as different as possible from one another. HCP2 subjects are unrelated “White - Not Hispanic” right-handed females aged 22-27. HCP3 subjects are “Black or African Am.” unrelated males aged 31-36, predominantly left-handed. HCP1 and HCP2 include 8 subjects for training and a separate 8 subjects for testing. HCP3 has only a training set of 8 subjects. Support vector machine classification, c.f. [27], of whole structural images from subjects in HCP2 and

HCP3 achieves around 90% accuracy on average over an 8-fold cross-validation test. Thus, brain structure strongly discriminates the two groups.

A further 8 training sets of 8 subjects (HCP-Random 1-8), randomly chosen from the remaining subjects in HCP Q3, enable tests of variability of performance over training sets.

HCP Lifespan data set

The HCP Lifespan Pilot Phase 1a, which is available online at <http://lifespan.humanconnectome.org>, has a much wider age range than the main HCP cohort including 26 subjects from 8-75 years. The HCP Lifespan diffusion-imaging protocol is a shortened version of the main HCP protocol with lower resolution (1.5 mm isotropic voxels) and only two HARDI shells, with $b = 1000$ and 2500 smm^{-2} . However, the protocol exploits the optimised features of the HCP scanners, providing data of considerably better quality than standard sequences.

Monkey data set

We also use 13 diffusion MRI data sets each from a separate perfusion-fixed vervet monkey (*Chlorocebus aethiops*) brain. The age range is 5-83 months. The animals were obtained from the Behavioral Science Foundation, St. Kitts and were socially housed in enriched environments. The experimental protocol was reviewed and approved by the Institutional Review Board of the Behavioral Science Foundation acting under the auspices of the Canadian Council on Animal Care. The brains were prepared using the protocol in [28]. They were imaged on a 4.7T Varian MR system at the Danish Research Centre for Magnetic Resonance,

Denmark, with imaging protocol similar to the 300mTm^{-1} protocol in [29], but with less directions per shell to enable full-brain imaging. Thus, each data set has three HARDI shells, with b -values approximately 2000, 3000, and 9500smm^{-2} , 16 $b = 0$ images, and isotropic 0.5 mm voxels.

Prisma data

We acquired two non-HCP in-vivo human data sets on separate subjects using the clinical 3T Siemens Prisma scanner in fMRIB, Oxford. The scanner is equipped with a modern gradient set of 80 mT/m maximum strength. The publicly available CMRR Multiband sequence [30] has been installed. These features allow acquisitions that resemble the main HCP diffusion MRI protocol. Subjects 1 and 2 are healthy male adults (29 and 33 years old respectively). Both data sets include a standard 3D T1-weighted MPRAGE (1mm isotropic resolution) and diffusion MRI data with three 90-direction HARDI shells, b -values of 1000, 2000, and 3000smm^{-2} , and 21 $b = 0$ images, each for two image resolutions: 2.5 mm and 1.35 mm isotropic voxels (TE=75 ms, TR=2.17 s, multiband factor=3 and TE=94.6 ms, TR=5.59 s, multiband factor=3, respectively). All data were acquired twice with reversed phase encoding direction (AP and PA). The HCP preprocessing pipelines [26] were used to correct for distortions in the diffusion MRI data and obtain transformations to T1 and MNI space. The Prisma scanner is less powerful than the bespoke HCP scanner and cannot achieve sufficient signal at 1.25 mm resolution, but the 1.35 mm data provides a pseudo ground-truth for IQT resolution enhancement of the 2.5 mm data.

Diffusion MRI models and fitting

We make use of four diffusion MRI models: i) the diffusion tensor (DT) model [22] and ii) Mean Apparent Propagator (MAP) MRI [31], to illustrate IQT resolution enhancement; and iii) NODDI [18] and iv) SMT [19], for IQT parameter mapping. Initial demonstrations of IQT resolution enhancement use DTI for simplicity. However, MAP-MRI provides a more general signal representation than DTI and can capture orientational variance in fibre crossing regions and across multiple b -values (for a fixed diffusion time), which IQT tractography requires. We demonstrate IQT parameter mapping by reconstructing NODDI and SMT maps from single-shell DTI maps.

We fit the DT model to only the $b = 0$ images and $b = 1000 \text{ s mm}^{-2}$ shell of the HCP and Lifespan data sets. Since the monkey data come from ex-vivo samples, the diffusivity of water in the tissue is 3–4 times lower [28, 32]. Thus, we use the $b = 0$ images and $b = 3000 \text{ s mm}^{-2}$ shell (87 gradient directions); the latter shows signal attenuation most similar to the $b = 1000 \text{ s mm}^{-2}$ HCP data. In all cases we use weighted linear least squares, cf. [33], accounting in the HCP data for the spatially varying b -value and gradient directions.

We fit the MAP basis up to order 4, giving 22 coefficients, by linear (unweighted) least squares to all three shells of the HCP data again accounting for spatial variation. The choice of scale parameters (see [31]) $u_x = u_y = u_z = 1.2 \times 10^{-3} \text{ mm}$ empirically minimises the error in fitting the signals in the HCP data.

The NODDI model, see [18] for details, has five free parameters: the intra-cellular volume fraction f_{ICVF} ; the free-water volume fraction f_{ISO} ; the mean (θ, ϕ) and variance parameter κ of a Watson distribution of fibre orientations. The orientation dispersion index (ODI) is a function of κ [18]. We fit the NODDI model via non-

linear fitting using the NODDI toolbox

https://www.nitrc.org/projects/noddi_toolbox version 0.9.

SMT [19] has two primary unknown parameters: the axial λ_{\parallel} and radial λ_{\perp} diffusivities of microscopic tissue compartments, which provide the microscopic FA, μFA . The framework also estimates the fibre orientation distribution p , which the scalar orientation dispersion entropy $H(p)$ summarises. We fit the model using the implementation from [19], which is available at <https://ekaden.github.io>.

IQT implementation

Our implementations of IQT use patch regression, i.e. learn mappings from a 3D neighbourhood, or “patch”, of N_1 voxels in the input (low-quality) data set to a corresponding patch of N_2 voxels in the output (high-quality) data set. Open-source code is available from the authors on request. Input and output voxels are vector-valued containing p_1 and p_2 values, respectively. The regression problem requires a training set of patch pairs $T = \{\mathbf{x}_i, \mathbf{y}_i\}_{i=1}^{|T|}$, where each input patch \mathbf{x}_i has dimension $p_1 N_1$ and each output patch \mathbf{y}_i has dimension $p_2 N_2$. Image reconstruction considers each patch of an input image independently; the learned mapping provides the output patch at the corresponding location. Thus, the IQT reconstruction step simply iterates over the image volume executing the mapping at each location to piece together the output image.

We consider a hierarchy of three types of mapping to implement the regression: global linear (for IQT-GL), regression trees, and regression forests (for IQT-RF).

Linear regression

Global linear regression computes a linear transformation matrix $G = YX^\dagger$, where the columns of Y are the training output-patches, \mathbf{y}_i , and the columns of X the corresponding input-patches, \mathbf{x}_i ; X^\dagger is an appropriate pseudo inverse of X . For an unseen input patch \mathbf{x} , the estimate of the corresponding output patch is $G\mathbf{x}$.

Regression trees

The regression tree implements a piecewise linear regression over the space of input data points [34, 35]. Each internal node in the tree sends data points into left or right subtrees by thresholding one of J scalar functions of \mathbf{x} , called *features*, F_1, \dots, F_J . Each leaf node contains a linear transformation with the same structure as the global linear transformation G defined above. Thus, for test input-patch \mathbf{x} , the output estimate is $G_t\mathbf{x}$ where G_t is the linear transformation associated with the leaf node t at which the data point arrives after traversing the tree.

Regression-tree training estimates the optimal i) tree structure, i.e. which nodes exist and which are internal and leaf nodes, ii) choice of feature and feature threshold at each internal node, and iii) linear transformation at each leaf node. We use a greedy search here similar to [35]. To control for overfitting, the algorithm reserves half the training set, T , as a validation set V . It selects features, thresholds, and fits linear models using the remaining half of T and accepts only splits that reduce the residual error of V . The procedure starts by assigning all training points to a single root node and computing a global linear transformation. It then seeks the feature-and-threshold combination that partitions the data with maximum information gain

$$I_0 - I_L - I_R, \tag{1}$$

where

$$I_0 = 2(|T| - |V|) \log \det(S) \quad (2)$$

is the information content of the root node, I_L and I_R that for the left and right child nodes, respectively, and

$$S = (Y - GX)^T (Y - GX). \quad (3)$$

For each candidate F_j , a standard golden search locates the threshold τ_j that maximises the information gain. We select the F_j with maximum information gain and accept the split if it reduces the residual errors on V . Specifically, the test compares the residual errors from the root node $\mathcal{E}_P = \sum_{i=1}^{|V|} \|\mathbf{y}_i - G\mathbf{x}_i\|$ with those from the children $\mathcal{E}_{LR} = \sum_{i=1}^{|V|} \|\mathbf{y}_i - C(G_L, G_R)\mathbf{x}_i\|$, where C selects transformation G_L in the left node if $F_j(\mathbf{x}) < \tau_j$ or G_R in the right node otherwise. We retain the split only if $\mathcal{E}_{LR} < \mathcal{E}_P$; otherwise, the root node remains a leaf node. If the split at the root node does reduce errors on V , the training process continues and recursively splits the training data into finer subsets assigned to left and right child nodes until the residual error of V stops reducing.

Regression forests

Regression forests [35] use multiple regression trees constructed from different bootstrap samples of the available training data. Outputs are element by element averages of the estimate from each tree weighted by the error covariance S_t^{-1} (c.f. Eq. 3) of the linear transformation G_t , estimated during training.

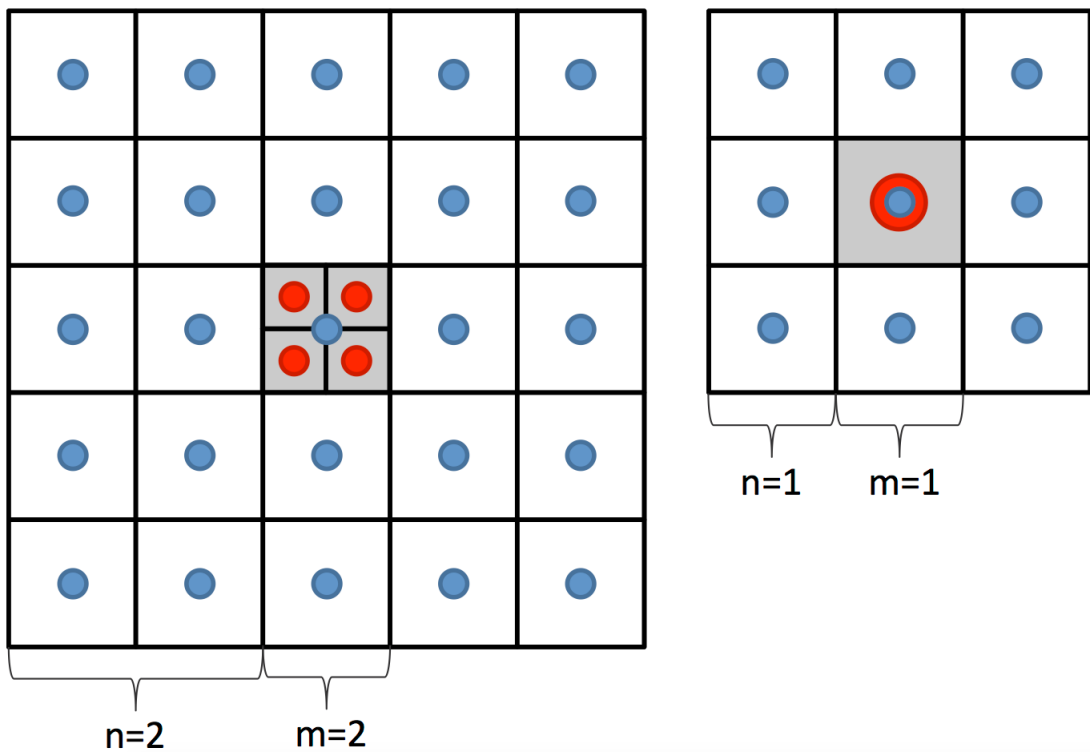


Figure 1: 2D illustration of the input (blue) and output (red) patch structure for $n = 2$ and $m = 2$ (typical for resolution enhancement) and $n = 1$ and $m = 1$ (typical for parameter mapping).

IQT resolution enhancement

The IQT resolution-enhancement mapping takes as input a $(2n + 1) \times (2n + 1) \times (2n + 1)$ cubic patch of voxels, so that $N_1 = (2n + 1)^3$, and outputs an $m \times m \times m$ cubic patch of voxels, so that $N_2 = m^3$. The output patch is a cubic array of subvoxels corresponding to the central voxel of the input patch, as figure 1 (left) illustrates. Each voxel in each patch contains $p_1 = p_2$ model parameters. For example, for DTI resolution enhancement with $n = 2$ and $m = 2$, each input vector, \mathbf{x} , contains the $p_1 = 6$ independent elements of the DT in each voxel of a $5 \times 5 \times 5$ low-resolution patch and each output vector, \mathbf{y} , contains the $p_2 = 6$ elements of each DT in the $2 \times 2 \times 2$ high resolution patch; the mapping is thus from \mathbb{R}^{750} to \mathbb{R}^{48} . Fourth-order MAP has $p_2 = 22$, leading to correspondingly higher dimensional mappings.

Training pairs come from downsampling each DWI by a factor of m in each dimension, fitting the model (e.g. DT or MAP) in each voxel of both the downsampled and full resolution image, and associating $(2n + 1)^3$ patches in the downsampled image with the m^3 patch in the full resolution image corresponding to the central voxel of the low-resolution patch. For downsampling, here we simply take the mean DWI intensity in each block of $m \times m \times m$ voxels, thus assuming an idealised point-spread function. The training data includes only patches wholly contained in the brain mask. We randomly subsample T from the pool of patch-pairs available from a set of training images without replacement. The HCP images contain on average around 7.8×10^5 brain voxels, which leads to a pool of around 3×10^6 patch pairs (allowing overlapping patches) with $m = 2$ and $n = 2$ in a training set of 8 HCP subjects. We consider sampling rates of up to 1 in 2, which hits the limits of available computer memory and use similar training set sizes from the Lifespan and Monkey data sets.

For IQT models trained on ex-vivo monkey data, a simple correction makes image patches from the in-vivo data sets compatible during reconstruction. Specifically, since the average eigenvalues of DTs fitted to the Monkey $b = 3000 \text{ s mm}^{-2}$ shell are approximately half of those from the HCP or Lifespan $b = 1000 \text{ s mm}^{-2}$ shell, we halve all the elements of each DT in the input patch, estimate the output patch from the regression model, and double all elements of the output to recover compatibility with the original data.

IQT parameter mapping

IQT parameter mapping estimates one set of image contrasts from another so, in general, $p_1 \neq p_2$. We assume here that the voxel grid of the output images is the same as the input, i.e. $N_2 = m = 1$, as in figure 1 (right). Thus to estimate NODDI or SMT parameters from DTI data, the mapping takes as input a $(2n + 1)^3$ cubic patch of DTs, so that $N_1 = (2n + 1)^3$ and $p_1 = 6$. For NODDI, the mapping outputs all five parameters at the central voxel of the input patch, so that $p_2 = 5$; for SMT it outputs λ_{\parallel} , λ_{\perp} , and $H(p)$ so $p_2 = 3$. Training pairs come from fitting DT maps to just the $b = 0$ and 1000 s mm^{-2} images (input) and matched NODDI or SMT maps fitted to full data sets (output).

Split features

For F_1, \dots, F_J , we use the following features of \mathbf{x} in IQT-RF to define partitions of the training data in the trees for both applications:

- The three eigenvalues of the DT in the central voxel.

- The linearity, planarity, isotropy [36] and trace of the DT in the central voxel.
- The means of each of the features above over the central $3 \times 3 \times 3$ cube (if input patch size $n \geq 1$ - see figure 1) and those over the whole $(2n + 1)^3$ cube (if $n \geq 2$).
- The orientational variance over the central $3 \times 3 \times 3$ cube (if $n \geq 1$) and the whole cube (if $n \geq 2$).

Thus here $J = 7$ if $n = 0$, $J = 15$ if $n = 1$, and $J = 23$ if $n \geq 2$. We quantify the orientational variance using the principal eigenvalue of the mean dyadic tensor, as in [37].

The set includes only orientationally and spatially invariant features to ensure generalisability among various types of training and test data. Additional features like absolute DT orientation and spatial location within the brain potentially enable the regression to build models specific to particular anatomic locations. Their inclusion leads to small reductions of error scores when training and testing within the HCP data set where the brain is consistently oriented within each image. However, they reduce the generalisability among data sets, because anatomical locations correspond less well between, say, monkey brains and HCP subjects than between different HCP subjects. Thus we exclude such features here.

Boundary effects

During IQT training, we use only patches wholly contained within the brain regions of the images. During testing, however, ignoring partial patches leaves a rim of unassigned values around the edge of the brain, which can erode a significant

portion of the image for large patches. We avoid this effect by completing partial patches with the conditional mean of all missing entries in \mathbf{x} , given the available entries, derived from a Gaussian model of the set of input patches in T . Specifically, we compute the mean $\bar{\mathbf{x}}$ and covariance $\Gamma = |T|^{-1}(X - \bar{X})^T(X - \bar{X})$ of the training input-patch library, where X is the matrix with columns \mathbf{x}_i , $i = 1, \dots, |T|$ and \bar{X} is the matrix of the same size as X , but with each column equal to $\bar{\mathbf{x}}$. These two quantities provide a multi-variate Gaussian model of the distribution of input patches. Equation 5.13 of [38] provides the conditional mean of a multivariate Gaussian distribution given a subset of components. We use that expression directly to replace the missing entries in boundary patches. Once completed in this way, we pass the completed \mathbf{x} through the IQT mapping in the usual way.

Tractography

Tractography experiments use the probabilistic tractography algorithm in [3] via the multi-shell multi-fibre reconstruction in [39]. IQT tractography simply runs the same algorithm after resolution enhancement with IQT. Specifically, we fit the MAP coefficients to low resolution DWIs and enhance resolution using IQT to obtain high resolution MAP coefficient maps. Those high resolution maps provide DWIs following the HCP protocol, through Eq. 24 in [31], which provides the input to the tractography package. IQT could work with any other tractography package in a similar way.

Error metrics

We use two simple error metrics to evaluate IQT performance. To compare DTI maps, we use the mean (over N_S test set subjects) median (over the set $\Omega(i)$) in the

i -th subject of brain voxels contained only in patches completely within the brain mask) root-mean-squared error (of the six independent DT elements D_1, \dots, D_6) against the gold standard (D_1^*, \dots, D_6^*):

$$\frac{1}{N_S} \sum_{i=1}^{N_S} \text{median}_{v \in \Omega(i)} \left(\sum_{j=1}^6 (D_j - D_j^*)^2 \right)^{\frac{1}{2}}. \quad (4)$$

We refer to this metric as DT-RMSE. We have also computed a range of other performance metrics, including orientational difference, as in [37], and normalised predicted signal error, as in [40], but all produce similar trends and conclusions so we present only DT-RMSE.

To compare NODDI maps, we use the mean (over subjects) of the mean (over NODDI parameters), median (over voxels) absolute error, which we refer to as NODDI-MAE. Only the three scalar NODDI parameters, f_{ICVF} , f_{ISO} , and κ , contribute and we normalise κ to $[0, 1]$, like f_{ICVF} and f_{ISO} .

Experiments and results

This section starts by demonstrating and evaluating IQT resolution enhancement, which we use to: a) highlight the advantage over standard operations such as interpolation, b) study the performance of IQT-GL and IQT-RF as a function of algorithm parameters, and c) demonstrate and evaluate generalisability beyond the training data. It then demonstrates the potential for parameter mapping from undersampled data sets using NODDI and SMT. We exemplify the downstream benefits of IQT resolution enhancement in tractography and, finally, verify results using independent test data (the Prisma data). Table 1 summarises the usage of the various data sets in each experiment and result.

Data set	Info	Figure number											
		2	3	4	5	6	7	8	9	10	11	12	13
HCP exemplar	1 subject	*	*	*				*				*	
HCP1 train	8 subjects	*	*	*	*	ab	ab	*	*	*	*	*	*
HCP1 test	8 subjects				*	ab	ab	*					
HCP2 train	8 subjects					b	ab						
HCP2 test	8 subjects						a						
HCP3 train	8 subjects						a						
HCP random 1-8	8 × 8 subjects						b						
Lifespan	26 subjects	*					b						*
Monkey	13 subjects	*					b						
Prisma	2 subjects								1	1	1,2		

Table 1: Data sets and their usage in each figure. The columns for figures 6 and 7 indicate which subfigure uses each data set. The entries in the Prisma data set row indicate which of the two subjects each figure uses.

Resolution enhancement

Figure 2 demonstrates IQT image-resolution enhancement and its generalisability. The test image (top left) is a 1.25 mm resolution DT map not in any training set. We show the direction-encoded colour (DEC) map [41], which summarises the full DT map with DT fractional anisotropy (FA) [42] as intensity and principal orientation as colour (red for left-right, green for anterior-posterior, blue for inferior-superior). The corresponding low-resolution image (top middle-left) comes from block averaging DWIs prior to model fitting to obtain a 3.75 mm resolution DT map (27 times larger voxels). The remaining DEC maps all show 1.25 mm resolution DT maps obtained from the 3.75 mm resolution map by interpolation or IQT. The IQT transformations operate on the low resolution DT maps; all use $m = 3$ and $n = 2$. The figure compares reconstructions from IQT-GL and IQT-RF both trained on HCP data (row 3 left) and IQT-RF trained on Lifespan and Monkey data (row 3 right). The IQT-GL transformation uses a training set of $|T| \approx 1.6 \times 10^6$ patch-pairs drawn from the HCP1 training set, as does each of the 8 trees in HCP-trained IQT-RF. Lifespan-trained IQT-RF also has 8 trees with similar $|T|$ to the HCP trees; Monkey-trained IQT-RF has 64 trees with $|T| \approx 0.2 \times 10^6$ so that the total number of training samples is similar to the HCP and Lifespan-trained forests. The choices correspond to the best operating points identified in later quantitative results (figure 6(a)); the optima are similar for $m = 2$ and $m = 3$. Linear and cubic interpolation is on the downsampled DWIs prior to model-fitting.

The difference maps in rows 2 and 4 are sums of absolute differences of DEC maps (before clipping) against ground truth (“Hi-res gold standard” in the figure). The bar chart (row 2 left) summarises later quantitative results (figures 5 and 7(b)) showing mean reconstruction errors (DT-RMSE) over the eight subjects in the

HCP1 test set.

IQT (e.g. third row middle-left) approximates the corresponding full-resolution image (top left) more closely than interpolation (e.g. top right). IQT provides sharper edges throughout the brain (e.g. pink arrows) than interpolation, producing less blurred images, and retains the high FA of the full resolution image, which interpolation artificially reduces. IQT also avoids artefactual “hotspots” (white arrows and pink arrows) that interpolation can cause [17]. The pink arrows highlight IQT’s ability to recover small structures obscured by partial volume effects at the resolution of the input data. Even IQT-GL improves substantially on interpolation, although the non-linear IQT-RF has additional benefits. For example, within homogeneous regions such as the corpus callosum (yellow ellipses), IQT-RF captures the smooth transition from red to yellow apparent in the “Hi-res gold standard”, whereas IQT-GL produces blockier output. The difference maps highlight better detail at tract interfaces (e.g. cyan circles) with IQT-RF than IQT-GL. In general, IQT avoids the strong anatomical structure that appears in the difference maps from interpolation. Even given only training data from the Lifespan and Monkey data sets, IQT-RF shows clear advantages over interpolation. In particular, the IQT-RF Monkey image (row 3 right) shows remarkable consistency with the IQT-RF HCP image (row 3 left-middle) given that the former is synthesised only from patches of ex-vivo fixed monkey-brain images.

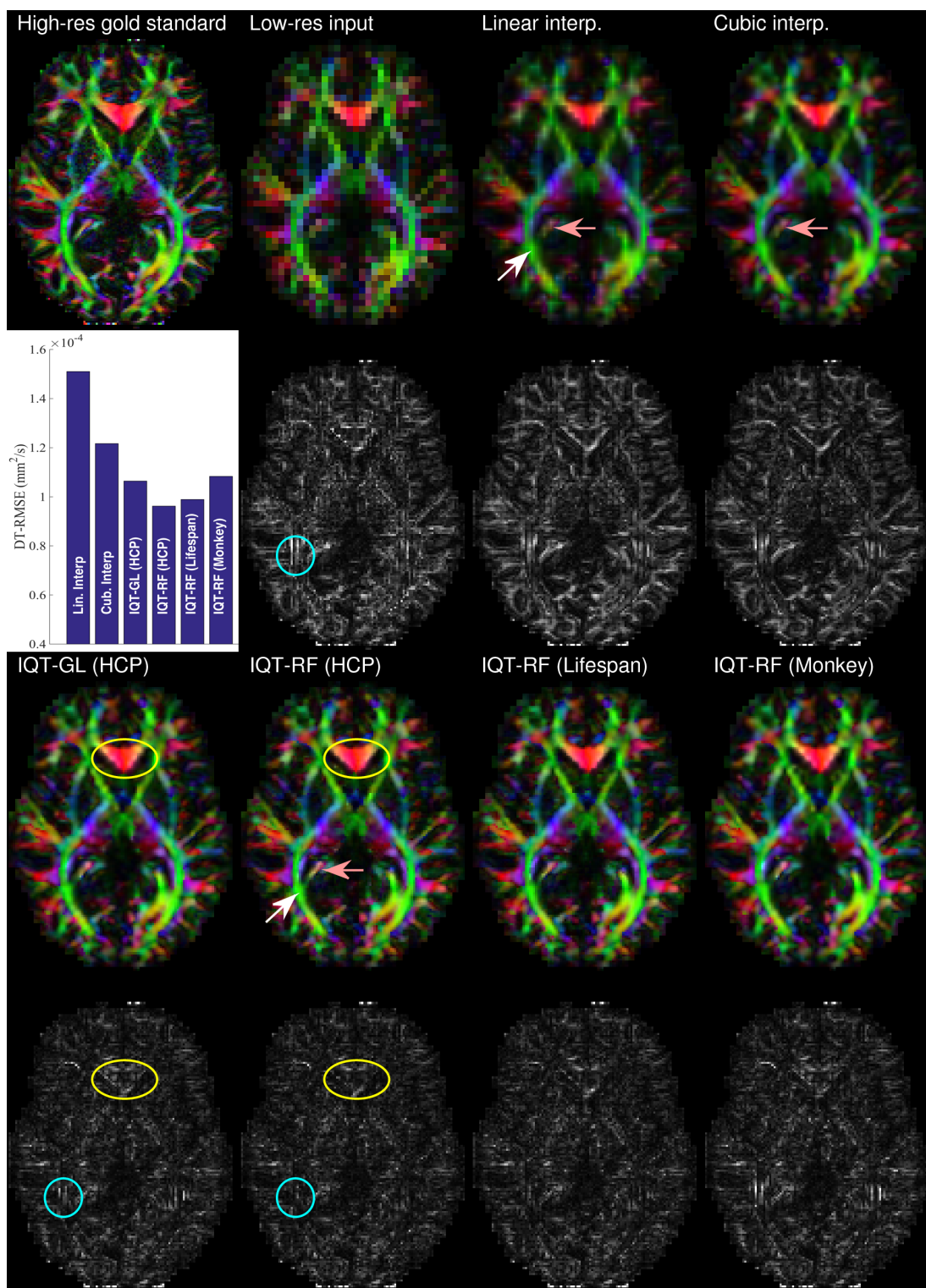


Figure 2: Demonstration of IQT image-resolution enhancement. The figure shows DEC maps (clipped at $FA=0.75$) for an example (not in any training set) full-resolution (1.25 mm isotropic) HCP DTI (top left), the corresponding image after downsampling the DWIs to 3.75 mm isotropic resolution (top left-middle), those after reconstruction to full resolution via linear (top right-middle) and cubic (top right) interpolation prior to DT fitting, and after reconstruction using IQT (third row) trained on other HCP data sets (left), the Lifespan data set (right-middle), and fixed monkey-brain data (right). The second and fourth rows show difference maps against the “Hi-res gold standard”. The bar chart on the second row compares average reconstruction errors over the 8 images in the HCP1 test set - lower is better.

Figure 3 provides a more detailed comparison of IQT and interpolation. It compares DTI principal direction maps in the cortex from a full-resolution HCP image (not in the training set) with those from the same image downsampled to 2.5 mm resolution and subsequently interpolated and IQT-RF reconstructed 1.25 mm images. A sharp transition in fibre orientation occurs across the boundary between grey matter and white matter (inner black contour) from tangential to orthogonal to the boundary. The transition is only visible with sufficiently high image resolution (see e.g. [28]) and in the figure is clear at 1.25 mm resolution but not at 2.5 mm (black arrows). IQT recovers the transition, as well as the directionality within the cortical ribbon, even though the low-resolution image contains little information in these areas, e.g. where indicated with black arrows. Interpolation fails to recover the transition as it cannot resolve the partial volume effects inherent in the low resolution input data, as discussed in [17]. In addition, IQT-RF recovers complex directional structure better in the white matter, particularly at interfaces, e.g. the area to the right of the white arrows.

Supplementary figure S1 shows an equivalent to Figure 3 without using boundary completion. The comparison reveals where boundary completion comes into play and confirms, qualitatively, efficacy of the approach. For example, at the crown of the gyrus containing the black arrow, the visual match of IQT-RF with the orig-

inal high resolution image is closer than that from interpolation. Supplementary figure S2 shows the random forest that provides the reconstruction in figure 3 to give some insight into its operation.

Parameter mapping

Figure 4 demonstrates IQT NODDI and SMT parameter mapping from single-shell data. IQT training uses the HCP1 training set with $n = 1$ and $m = 1$ (best operating point from figure 6(b)), $|T| = 1.6 \times 10^6$, and 8 trees. The second column comes from inputting just the $b = 0$ and 1000 s mm^{-2} images to the standard NODDI and SMT estimation tools, as used for the first column. The SMT algorithm uses random starting points for single-shell input.

As [18, 19] predict, conventional estimation techniques (“Direct Fitting” in the figure) fail given single-shell input, producing strongly disrupted maps, e.g. noisy appearance of f_{ICVF} and μFA . IQT-GL also performs poorly and fails to recover the structure of most parameter maps. The improvement IQT-RF provides is striking. It recovers the broad patterns of contrast between grey matter, white matter, and cerebro-spinal fluid in all parameters, albeit with less point-to-point variation (cyan ellipses) and weaker contrast (yellow ellipses) than the “Original”.

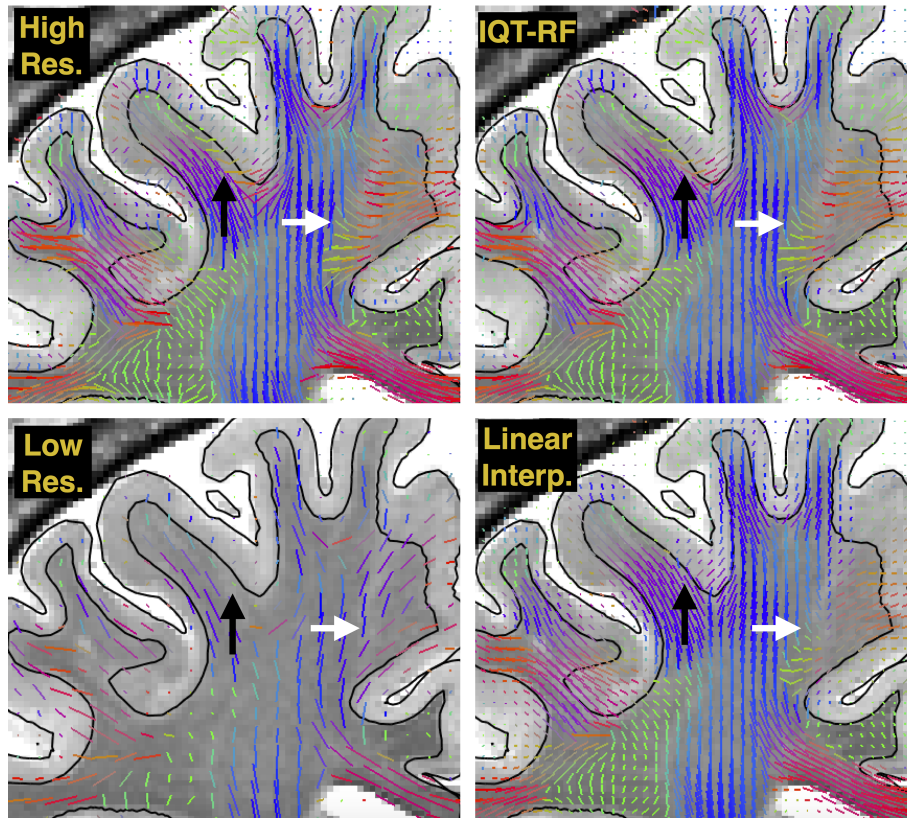


Figure 3: Visualisation of DTI principal directions in and around the cortex in one HCP data set: original 1.25 mm resolution image (top left), downsampled 2.5 mm resolution image (bottom left), IQT-RF ($m = 2, n = 2$) reconstructed 1.25 mm resolution image (top right), and linearly interpolated 1.25 mm resolution image (bottom right). The two black lines correspond to the pial and white matter boundary surfaces so encompass the cortical ribbon. The background is a high-resolution T2-weighted image. Each vector has been scaled by the corresponding FA and coloured by orientation in the same way as the DEC maps in figure 2.

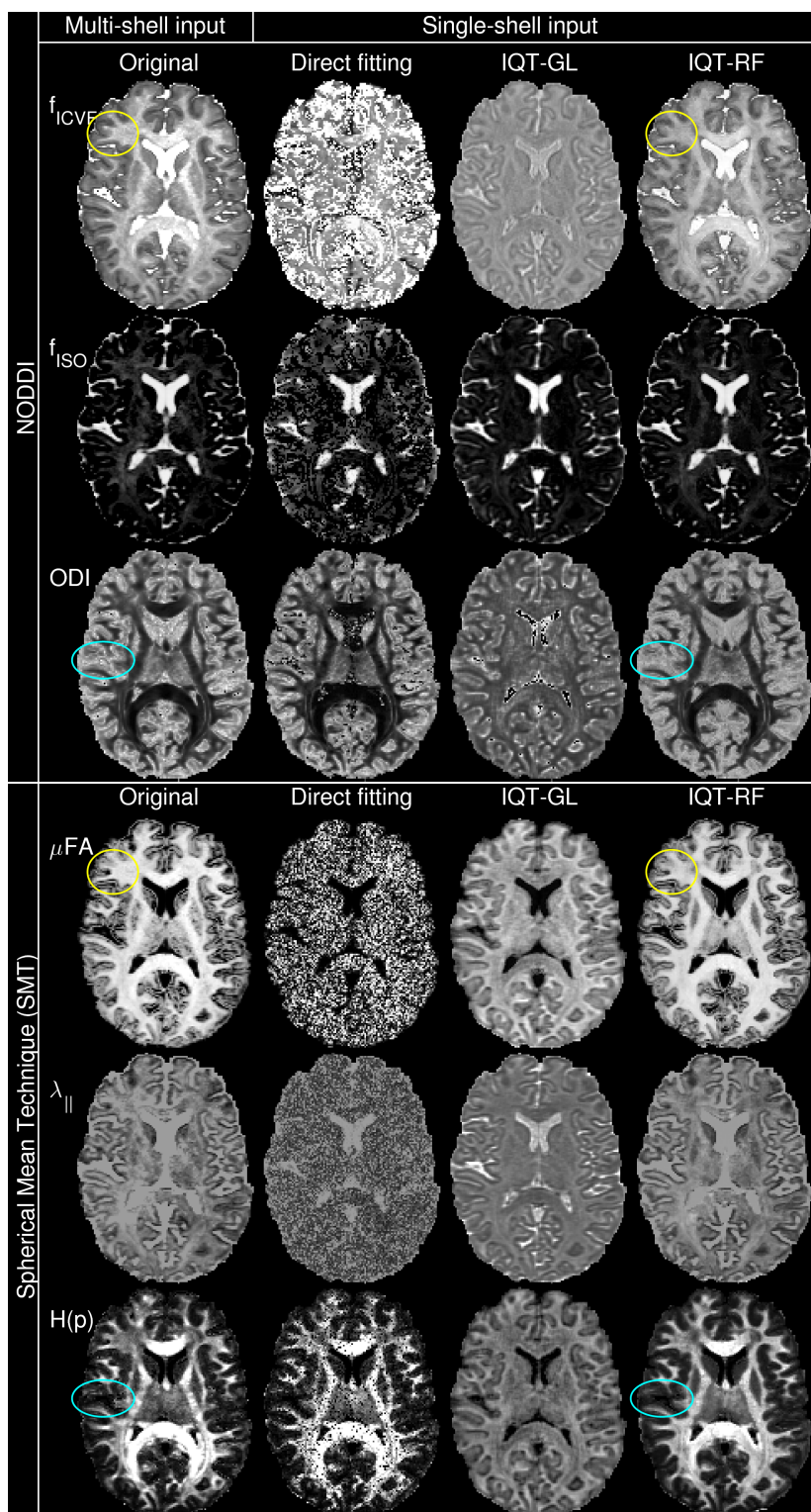


Figure 4: IQT reconstruction of NODDI and SMT parameters from DTI; trained on HCP1. The left column shows ground truth NODDI (top three rows) and SMT (bottom three rows) parameter maps estimated using standard tools from the full data set (all 3 non-zero b -value shells) of the same HCP subject as figure 2. The second column shows corresponding maps estimated using standard tools from just the $b = 0$ images and $b = 1000 \text{ s mm}^{-2}$ shell. The third and fourth columns show maps recovered by IQT-GL and IQT-RF, respectively, from just the $b = 1000 \text{ s mm}^{-2}$ DTI.

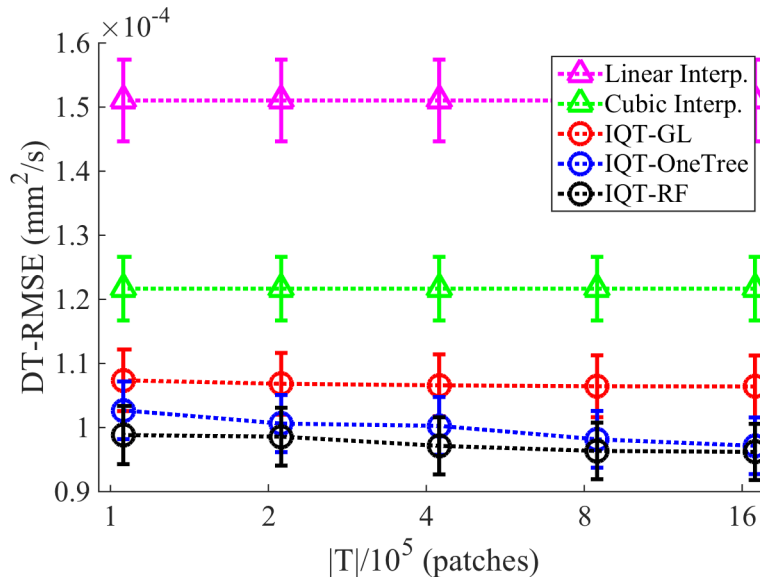


Figure 5: High-resolution DTI reconstruction errors as a function of per-tree training set size, $|T|$, for interpolation and IQT, using HCP1 training and test sets with $m = 2$ for all algorithms, and $n = 2$ for IQT. The error bars show one standard deviation over subjects.

Quantitative evaluation

Figure 5 compares high-resolution DTI reconstruction errors (DT-RMSE from Eq. 4) on the HCP1 test set for various algorithms as a function of per-tree training set size, $|T|$. As in figures 2 and 3, low-resolution input data comes from block averaging DWIs prior to model fitting and linear and cubic interpolation is on the downsampled DWIs prior to model-fitting. All quantitative results on resolution enhancement, in this and subsequent figures, use $m = 2$ (factor of 8 difference

in voxel volume), which is representative of practical situations. However, this is in contrast to the qualitative results in figure 2, which use $m = 3$ (factor of 27 voxel volume difference) to emphasise visible differences in reconstruction quality better. In the IQT-RF data points, in all figures in this section, the number of trees is inversely proportional to the number of training samples per tree so that the total amount of training data is comparable; from left to right, the number of trees is 64, 32, 16, 8, 4.

Figure 5 shows that standard-deviation over subjects (error bars in the figure) is consistent among the different algorithms. However, the subject-specific error score is highly correlated among algorithms, because individual anatomical features, such as the size of the ventricles, strongly drive its absolute value. For this reason, standard deviation over training sets is more meaningful when comparing performance between algorithms than standard deviation over subjects. Later results (figure 7(b)) show that the standard deviation over training sets has a value of around $10^{-7} \text{ mm}^2\text{s}^{-1}$, which is negligible on the scale of figure 5. This suggests that even small differences in error score, such as between IQT-OneTree (a single regression-tree) and IQT-RF at $|T| \approx 1.6 \times 10^6$, are strongly significant at about 60 standard deviations. Similarly, the differences in performance between interpolation and IQT-GL, or IQT-GL and IQT-RF, are highly significant. Thus we conclude that, as the qualitative results in figure 2 suggest, all IQT implementations outperform linear and cubic interpolation, and non-linear random forests (IQT-RF) outperform non-linear decision trees (IQT-OneTree), which outperform global-linear regression (IQT-GL). The flexibility of the non-linear mappings captures significant extra detail. Moreover, as is common in learning applications, the consensus approach of the forest provides subtle but significant additional benefit over the single tree; this is both because it enables the learning to consider more training data, and because it mitigates the greedy tree-learning algorithm hitting

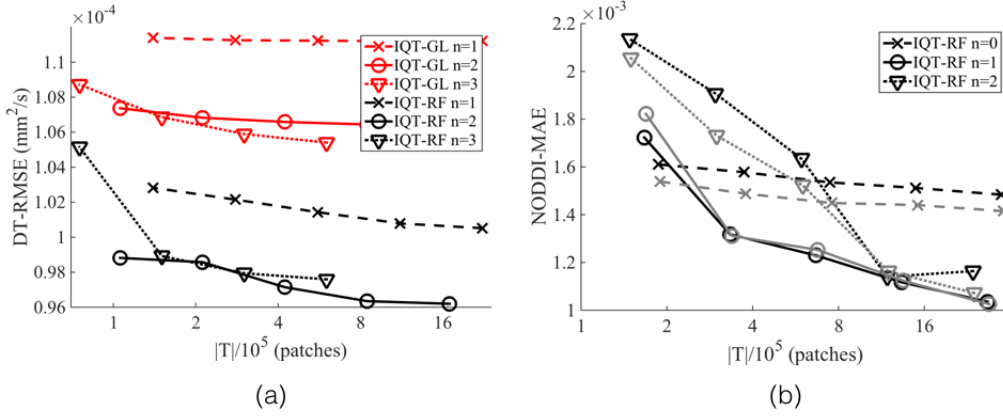


Figure 6: Reconstruction errors as a function of training set size, $|T|$, and patch size, n . Panel (a) is for IQT resolution enhancement with $m = 2$ using HCP1 training and test sets. Panel (b) is for IQT NODDI parameter estimation. The black markers and lines show error scores for models trained on HCP1 and grey those for HCP2 training; the test set in both cases is from HCP1.

local minima.

Figure 5 also shows that while IQT-GL benefits little from increasing $|T|$, and interpolation is obviously independent of $|T|$, IQT-RF performance increases steadily with $|T|$. Thus, IQT-RF favours a small number of complex trees, from large $|T|$, over a large number of smaller trees trained with less data.

Figure 6 plots reconstruction errors against $|T|$ for various n to guide the choice of the parameter. Panel (a) is for IQT resolution enhancement and (b) for NODDI parameter mapping. For comparison to the minimum NODDI-MAE score for IQT-RF of 0.10×10^{-2} ($n = 1$, maximum $|T|$), the corresponding score for “Direct Fitting” (figure 4) is 1.81×10^{-2} (18 times worse than IQT-RF), and the best case for IQT-GL ($n = 2$, largest $|T|$) is 0.65×10^{-2} (6 times worse than IQT-RF). We omit error bars in both panels for clarity, but they have similar magnitude to those in figure 5.

In resolution enhancement, performance of IQT-GL increases with n for sufficient

$|T|$. This reflects the increase in linear-model complexity as n increases: for $n = 1$, the input patch includes 27 voxels with $m = 2$, so the linear transformation has $7824 = 163 \times 48$ elements ($163 = (2n + 1)^3 \times p_1 + 1$; $48 = m^3 \times p_2$; $p_1 = p_2 = 6$); the number of parameters increases to 98832 ($= (7^3 \times 6 + 1) \times (2^3 \times 6)$) for $n = 3$. In contrast, the random forests find an appropriate level of complexity of the model for any n . However, computer memory limitations enforce a trade-off between number of data points the training set can contain (maximum at small n) and the information content of each data point (maximum at large n). Figure 6(a) shows that $n = 2$ optimises that trade off.

Supplementary results in figure S3 explore the dependence of performance on noise level in the input data and orientation with respect to the training data (representing a change in head orientation). They show that performance degrades gracefully as noise increases and that performance is orientation dependent. Other results (not shown) reveal that reconstruction error reduces rapidly as the number of trees in the random forest increases from 1 to 3, but stabilises above 4; increasing the number of trees further, even for high $|T|$, provides little gain. Moreover, for fixed training set size, little improvement arises from increasing the number of source images in the HCP1 training set above its setting here of 8. Corresponding results for MAP-MRI resolution enhancement (not shown) show similar trends to those for DTI: forests outperform single trees, global linear transformations, and standard interpolation in terms of reconstruction error; $n = 2$ is a good choice, although for MAP-MRI $n = 1$ performs similarly to $n = 2$ and we thus use $n = 1$ for MAP-MRI driven tractography later, as the boundary completion is faster to run.

Figure 6(b) shows that $n = 1$ optimises the neighbourhood size for IQT NODDI parameter estimation. The comparison to $n = 0$ confirms that the neighbourhood

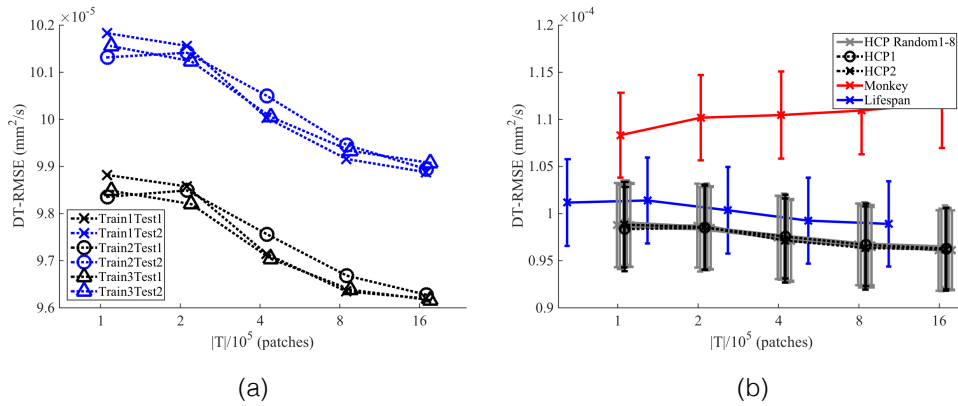


Figure 7: Quantitative evaluation of generalisability of IQT resolution enhancement both within the HCP data set (panel (a)) and across different data sets (panel (b)). Panel (a) plots errors for training and test sets of demographically distinct HCP sub-groups with $m = 2$ and $n = 2$. The same colours indicate the same test set (HCP1 or HCP2) and the same markers indicate the same training set (HCP1, HCP2, or HCP3). HCP1 contains diverse subjects while HCP2 and HCP3 are homogeneous but distinct. In the legend, “TrainXTestY” represents training on HCPX and testing on HCPY. Panel (b) shows reconstruction errors on the HCP1 test set for IQT-RF with $m = 2$ and $n = 2$ trained on Monkey data, Lifespan data, the HCP1 and HCP2 training sets, as well as the eight HCP-Random training sets. The error bars show one standard deviation over subjects.

information is important and improves substantially on a direct estimation of the NODDI parameters from each DT in isolation.

Figure 7 evaluates generalisability of IQT beyond its training set. Panel (a) evaluates the effect of choice of training set within the HCP data set, while (b) compares the performance of models trained on HCP, Lifespan, and Monkey data sets.

Figure 7(a) tests the hypothesis that IQT models generalise well across demographic groups, i.e. performance depends little on the choice of subjects in the training set. Results show that reconstruction errors on a test set of images from a very homogeneous group (HCP2: white, right-handed, females in their 20s) depend very little on whether the training set comes from other individuals in the same group, a set of individuals in a distinct homogeneous group (HCP3: black,

left-handed, males in their 30s), or a diverse group (HCP1: with all combinations of the characteristics in HCP2 and HCP3). In fact, training on HCP1 produces slightly lower errors than training on HCP2 on both test sets, highlighting that a precise match of demographics between the training and test sets has little importance. Error scores vary substantially among test sets, reflecting the relatively high variability over subjects, which, as discussed earlier, depends on gross anatomy, such as ventricle volume.

Figure 7(b) tests wider generalisability of IQT-RF by comparing DTI reconstruction errors for the HCP1 test set using IQT-RF trained on various different data sets. The variability of the error scores among the HCP Random 1-8 training sets (grey markers) allows us to evaluate the standard deviation over training set, which is of order $10^{-7} \text{ mm}^2\text{s}^{-1}$. This illustrates the within-subject variability (in contrast to the between-subject variability that the standard deviation over subjects reflects) that we refer to earlier in the discussion of figure 5 and confirms that performance depends only weakly on the choice of HCP training subjects. Although error scores from the Lifespan and Monkey data sets are outside the range of variability of the HCP training sets, so significantly worse statistically, both training sets still produce good approximations to the ground truth. Both produce much lower error scores than interpolation techniques: comparison of error scores with figure 5 shows that IQT-RF trained on Monkey data produces similar error scores to IQT-GL trained on HCP data (see also figure 2 inset).

Figure 7(b) also shows that for the Lifespan training sets, performance increases with $|T|$ favouring a small number of complex trees, as with the HCP training sets. However, for the Monkey data sets, a larger number of simpler trees proves favourable; this promotes generalisability better for this training data, which is more distinct from the test data than the HCP or Lifespan training data.

Tractography

Figure 8 demonstrates IQT’s direct benefits to tractography in ameliorating ambiguities that arise from limited spatial resolution. The top row replicates the finding in [3] that probabilistic tractography separates the four distinct pathways from the cortical hand area to the thalamus, brainstem, spinal cord and putamen [43] when using 1.25 mm resolution data sets. Figure 4 of [3] also shows that tractography cannot separate the pathways using 1.5 mm or 2 mm resolution images. However, following IQT MAP-MRI resolution enhancement ($n = 1$, $m = 2$, HCP1 training data, $|T| \approx 1.6 \times 10^6$, 8 trees in IQT-RF), probabilistic tractography recovers the four pathways even from 2.5 mm resolution input data. The finding is consistent over the HCP test subjects. In contrast, the same tractography procedure after linear or cubic interpolation misses at least one of the four pathways and produces various false positives (yellow arrows).

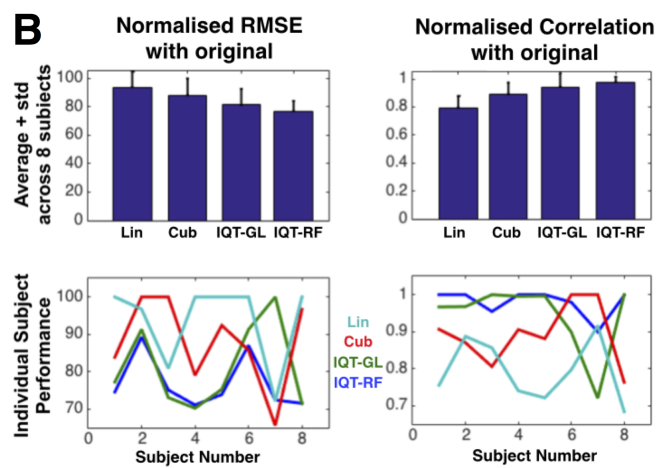
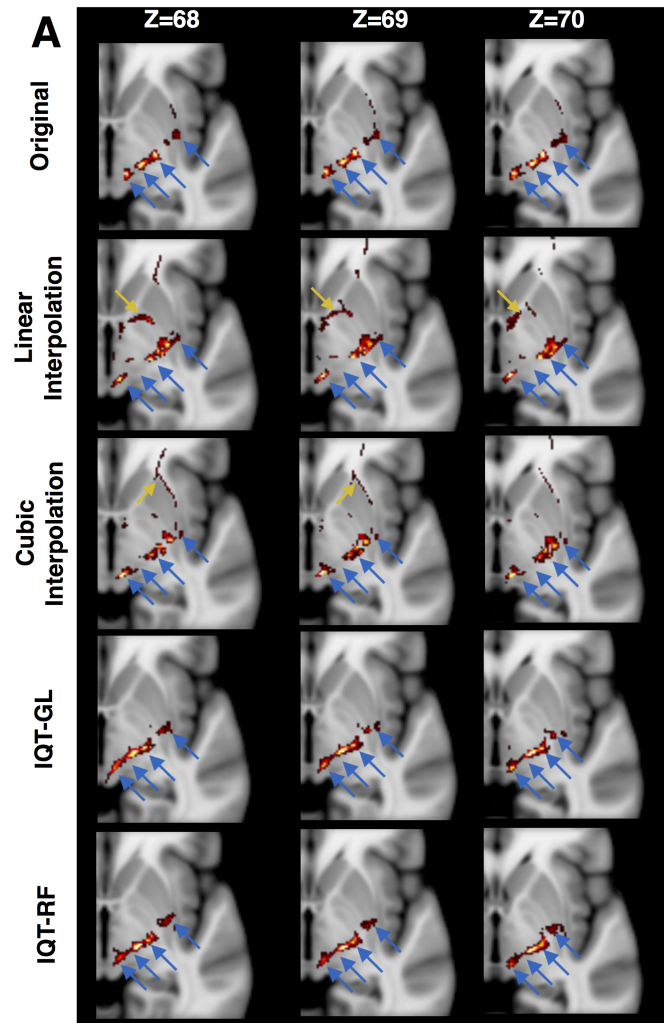


Figure 8: IQT-enhanced tractography. A: Probabilistic tractography maps (thresholded at 0.35%) for one HCP subject (not in any training set) overlaid on a group-mean T1-weighted image. Tractography maps come from the original high resolution data set with seed region in the hand area of the motor cortex, together with corresponding maps using data sets reconstructed via linear and cubic interpolation, IQT-GL, and IQT-RF, after downsampling to 2.5 mm resolution (8 times larger voxels). The blue arrows point to four cortical projections that tractography identifies in the original data (from medial to lateral: cortico-thalamic, cortico-bulbar, cortico-spinal and cortico-striatal). B: Error and correlation scores between probabilistic tractography maps from the original high resolution data set and each reconstruction; bar-chart metrics are averages over the 8 test subjects from HCP1. Individual performance on all test subjects are root-mean-square errors and correlations normalised to the worst and best case, respectively, so the worst error score is 100 and the best correlation score is 1.

IQT on independent data sets

Figures 9, 10 and 11 demonstrate IQT-RF, trained on HCP data, on the independently acquired Prisma data sets. The key distinction with earlier results in figures 2 and 3, is that the low resolution data set is acquired at 2.5 mm resolution rather than obtained by the same downsampling procedure used in training. The results confirm efficacy of IQT-RF on realistic data. Specifically, IQT-RF sharpens weak structures in data from widely available scanners (figure 9), produces plausible NODDI parameter maps even from very sparse input data sets (figure 9), recovers structure at the white-matter-grey-matter boundary that interpolation cannot (figure 10), and enables tractography to separate thin pathways from data acquired at 2.5 mm isotropic resolution (figure 11).

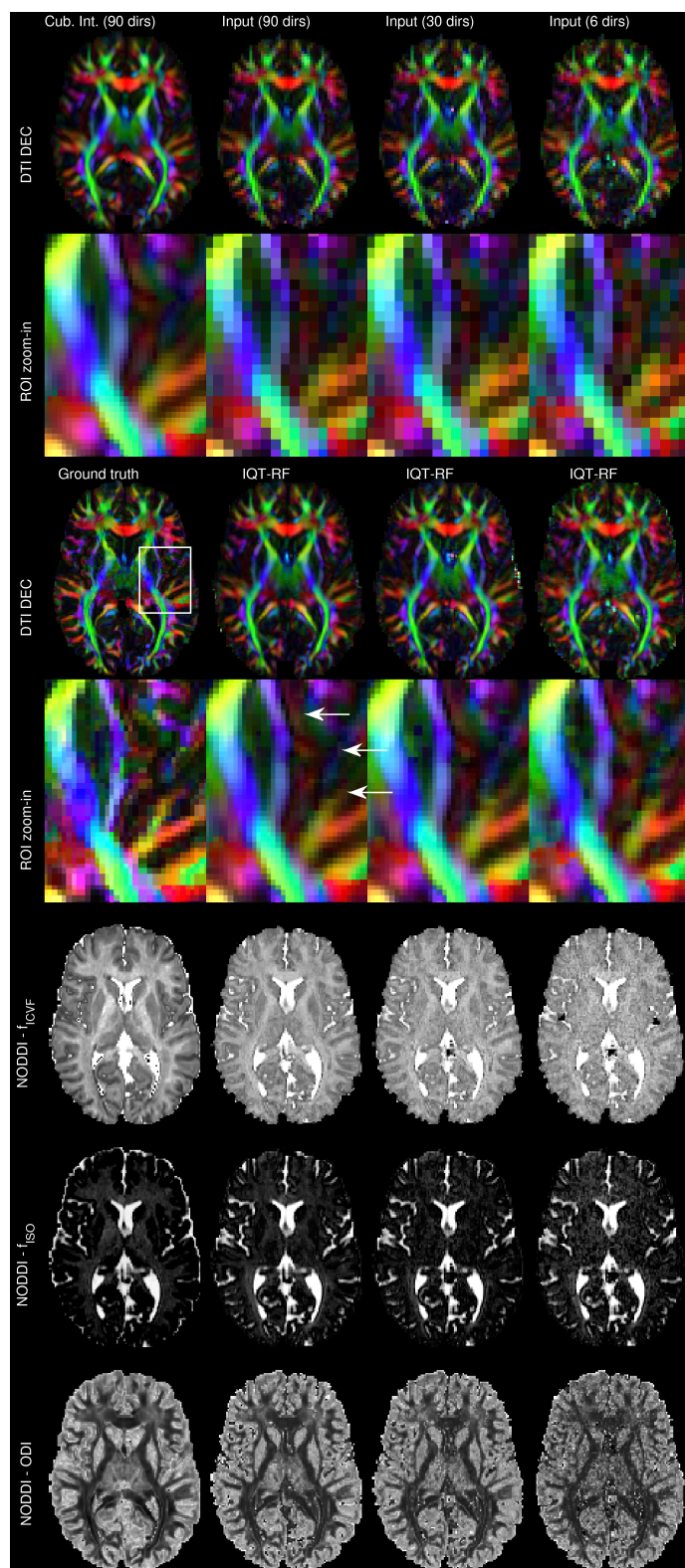


Figure 9: IQT-RF for the non-HCP data set from Subject 1. The top row (columns 2-4) shows DEC maps from the $b = 1000 \text{ s mm}^{-2}$ shell of the 2.5 mm data set with all 90 directions (middle left), a subset of 30 of the 90 directions (middle right), and a subset of 6 directions (right). The top left panel shows the DEC map after cubic interpolation of the 90 DWIs used in the middle-left panel to 1.25 mm resolution. The third row compares DEC maps from the 1.35 mm pseudo ground-truth data set (left) with those from 1.25 mm DTI maps reconstructed by IQT-RF from the 2.5 mm DTI maps with varying numbers of input images in the corresponding columns of the top row. To show differences more clearly, rows 2 and 4 zoom in on the area marked by the white rectangle in the left-most image of row 3 in the corresponding image immediately above; the white arrows highlight structures visible in the IQT maps that are not discernible in the low-resolution images. The last three rows compare ground-truth NODDI parameter maps (left), fitted to the whole 1.35 mm data set, to parameter maps reconstructed by IQT from DTI fitted to 90 (and 21 $b = 0$ images), 30 (plus 5 $b = 0$), and 6 (plus 1 $b = 0$) (middle left to right, respectively) directions from the $b = 1000 \text{ s mm}^{-2}$ shell of the 1.35 mm data set. In each case, IQT-RF uses an eight-tree random forest learned from HCP1 with $|T| \approx 1.6 \times 10^6$; $n = 2$ for DTI resolution enhancement, and $n = 1$ for NODDI mapping.

Figure 9 illustrates DTI resolution enhancement and NODDI parameter mapping via IQT-RF on the Prisma data set from Subject 1. The increase in detail in the high resolution images obtained from IQT-RF is clear in many areas of the DEC maps, particularly after zooming in (rows 2 and 4). For example, the small red, blue, and yellow (top to bottom) structures marked with white arrows are clearly discernible in the IQT-RF maps, but not in the low-resolution 2.5 mm data, even after cubic interpolation of the source DWIs to 1.25 mm resolution. The figure also shows that performance degrades gracefully as the number of input DWIs decreases. The second column of the figure uses all 90 $b = 1000 \text{ s mm}^{-2}$ DWIs to construct the DTI input to IQT-RF. Such a data set would require about 15 mins to acquire on a modern clinical system. However, the third and fourth columns use only 30 and 6 DWIs, respectively, representing sparser 5 minute or 1 minute acquisitions. Very short acquisitions like these are common in clinical situations where subjects cannot stay still, e.g. infants [44], or are in the middle of a surgical operation [45]. The figure shows that even using very sparse input to IQT,

the mappings we learn produce high resolution images with good consistency to those from less noisy input and to the pseudo-ground-truth. IQT-RF also produces NODDI maps that reflect the ground-truth contrast. IQT-RF preserves the pattern of contrast even with the thirty-direction input, although the maps become visibly disrupted with six-direction input.

Figure 10 compares cortical DTI principal direction maps in the 1.35 mm and 2.5 mm resolution data sets with those after interpolation and IQT-RF resolution-enhancement from 2.5 mm to 1.25 mm resolution. The figure illustrates areas where IQT-RF recovers the sharp tangential-orthogonal transition at the white-grey matter boundary that interpolation loses (black arrows and ellipse). It also shows areas that match the 1.35 mm pseudo ground-truth less well (yellow arrow and grey matter above black ellipse). In the white matter of the gyrus indicated by the yellow arrow, IQT-RF and interpolation both produce a slightly different orientation (more purple) to the 1.35 mm data (more green). However, the orientations in both reconstructions reflect the source 2.5 mm data, so the difference may arise from imperfect spatial alignment of the 1.35 mm and 2.5 mm data sets; small residual misalignments always remain even though state-of-the-art image registration [46] maps both data sets to MNI space. IQT-RF recovers the sharp tangential-orthogonal transition across the white-grey matter boundary on the left side of that sulcus, albeit displaced (again possibly arising from imperfect spatial alignment with the pseudo-ground truth), whereas interpolation does not recover the transition at all. The grey matter at the top of the sulcus (right of yellow arrow, above black ellipse) has complex structure in the 1.35 mm data set that is lost in the 2.5 mm data; neither interpolation nor IQT-RF recover the full complexity of the region. IQT-RF relies on boundary completion in much of that area. Nevertheless, reassuringly, IQT-RF remains consistent with the low-resolution data producing similar output to interpolation.

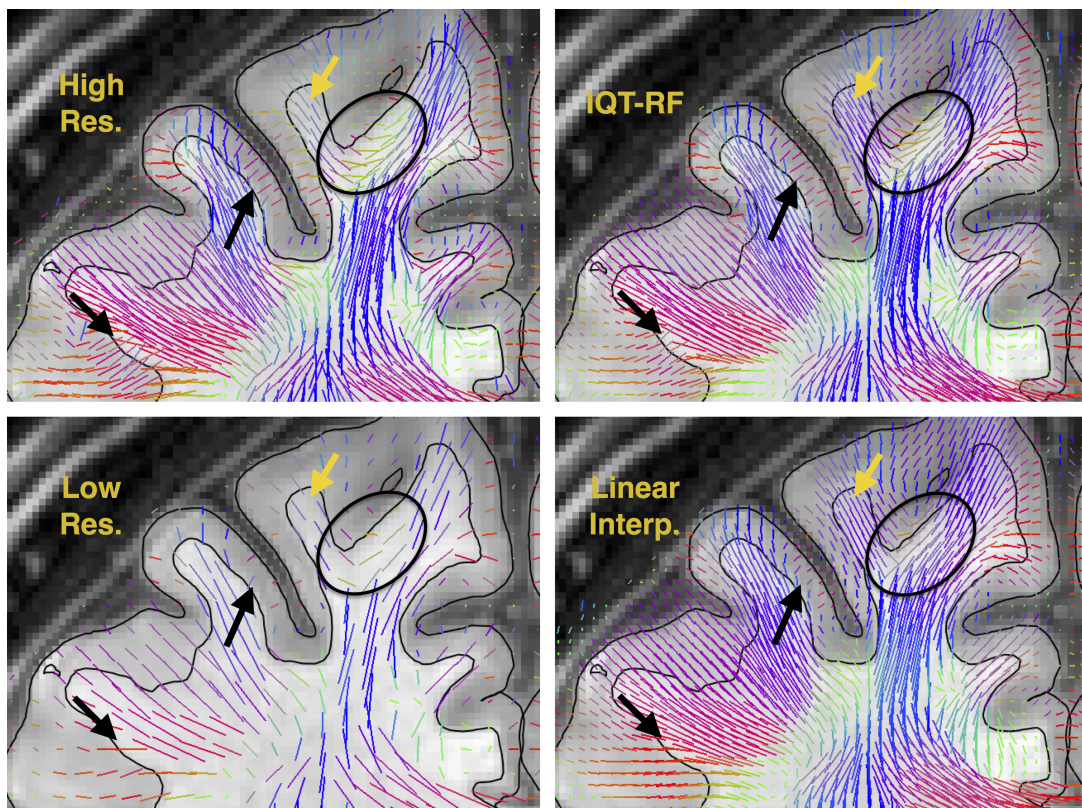


Figure 10: Visualisation of DTI principal directions in and around the cortex in Prisma data, Subject 1, before and after resolution enhancement.

Figure 11 shows probabilistic tractography from the hand area of the motor cortex using the full 1.35 mm data set, and 1.25 mm data sets reconstructed by interpolation and IQT-RF MAP-MRI resolution enhancement ($n = 1$, $m = 2$) from the 2.5 mm Prisma data sets using all 90 directions.

In both subjects, tractography separates the four cortical projections (as described in Figure 8 and indicated by the blue arrows) from the pseudo ground-truth 1.35 mm data set. After interpolation of the 2.5 mm data set to 1.25 mm resolution, tractography cannot separate the four pathways, in particular missing the cortico-bulbar projection to the brainstem (2nd projection from the left), and produces several false positives. Tractography on the 1.25 mm IQT-RF data sets finds four pathways, recovering the cortico-bulbar projection. In subject 2, tractography on the IQT-RF 1.25 mm data set separates the three medial pathways even more clearly than the 1.35 mm data set. In both subjects, the position and shape of the most lateral (cortico-striatal) projection is somewhat disrupted in comparison to the 1.35 mm data set. However, the match between the IQT-RF tractography and the pseudo ground-truth is remarkable given that the former uses less than one sixth of the input data.

Discussion

We have introduced the notion of IQT, proposed two straightforward implementations, and demonstrated the potential to transfer information from a unique custom-built high-power imaging system to lower-quality data representative of widely available machines. The current implementations of IQT require only off-the-shelf machine learning tools and simple and standard data preprocessing. They require no image smoothing, denoising, spatial normalisation, or specialised

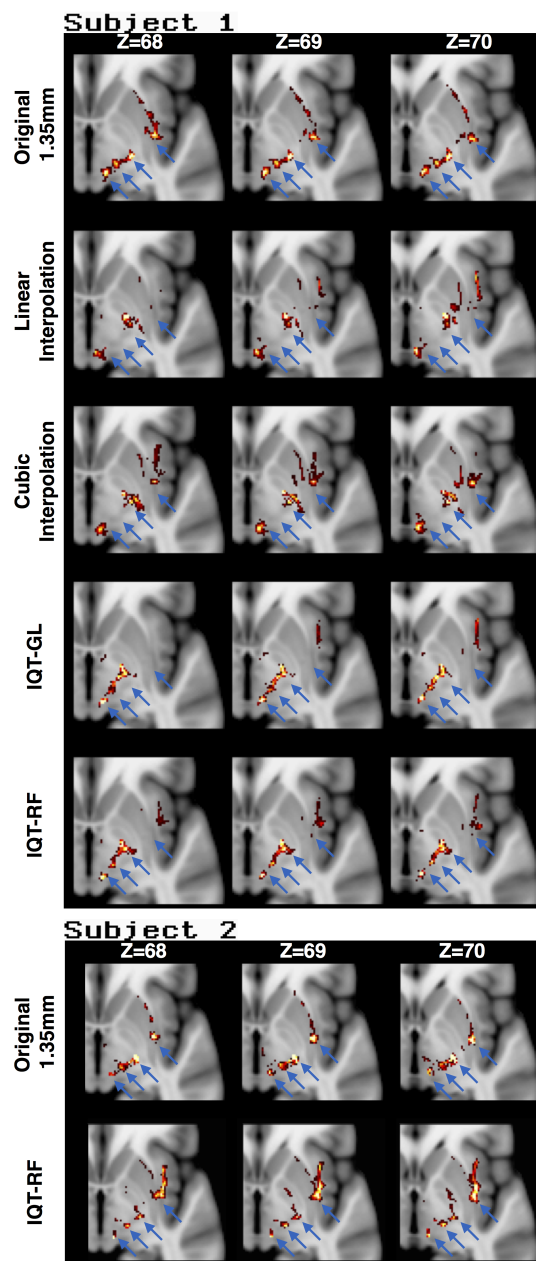


Figure 11: Probabilistic tractography maps, overlaid on T1 maps in MNI space, on both subjects for various acquired and reconstructed data sets.

acquisition sequence or protocol. Although training IQT-RF can require days of processor time, a straightforward image reconstruction with IQT-RF typically requires about 10 minutes of processor time (similar to DTI or MAP-MRI processing, but $< 5\%$ of the standard NODDI computation), since it requires just one linear transformation per image patch. Boundary completion increases computation. However, both operations also parallelise naturally.

IQT resolution enhancement shows clear improvements over interpolation, which is the standard way to estimate sub-voxel information in imaging applications from visualisation by a radiologist to sophisticated post-processing procedures, such as tractography. IQT avoids artefacts commonly associated with interpolation [17], such as hot-spots/blurring in figure 2 and the partial volume effects in figure 3, and recovers fine detail lost at low resolution. Although higher order interpolation can also reduce such artefacts to some extent [17], such methods all approximate high resolution voxels as averages over low-resolution neighbourhoods. IQT’s strategy is fundamentally different and has the substantial advantage of exploiting knowledge of likely corresponding high resolution image structure. The technique should be beneficial in any application that requires sub-voxel inference.

Future work must compare performance of IQT’s mappings learned from training data, with more recent interpolation techniques, such as sparse reconstruction or non-local means [6, 47]. Preliminary tests with the basic algorithm in [47] find best performance running it directly on the DT element maps, which gives DT-RMSE of $1.1 \times 10^{-4} \text{ mm}^2 \text{ s}^{-1}$ on the HCP1 test set. Thus, it scores in between cubic interpolation and IQT-GL; see bar chart in figure 2. The later technique [6] tailored specifically for DWI will be interesting to compare once the code is freely available.

The generalisability that figures 2 and 7 illustrate is important, because it reveals the potential for IQT to enhance images distinct from the training set. Specifically, reconstruction of high-resolution DTI degrades remarkably little as we move through IQT trained on i) HCP data from very similar subjects to the test data; ii) HCP data from a distinct demographic; iii) data from a distinct acquisition protocol and very different demographic (the HCP Lifespan data set); and iv) data from fixed monkey brains. This shows the great potential of IQT to enhance any data set even when training data is not directly representative. Nevertheless, the robustness of performance in the presence of pathology remains a key question for healthcare applications. Future work must determine whether IQT can benefit cohort studies and whether training data from patients is necessary to do so. A related challenge is to quantify uncertainty in IQT output, to highlight areas of low confidence e.g. from lack of representation in the training data. The current implementations naturally provide such estimates, via Eq. 3, although future work must evaluate their precision. See [48] for preliminary work on IQT uncertainty quantification in the presence of pathology. More generally, future work could explore ideas from domain adaptation [49] to address the challenges of generalisability particularly in situations where direct training data is hard to come by.

One natural question for future work on resolution-enhancement is whether IQT can predict even higher resolution images than its training data on the assumption of fractal properties of brain images, i.e. self-similarity across scale. This is reasonable over a limited range of scales, as, for example, we can expect similar partial volume effects to arise going from 1.25 mm to 2.5 mm voxels as from 0.625 mm to 1.25 mm. Figure 12 shows an initial exploration of the idea. It shows 0.625 mm and 0.313 mm isotropic resolution DEC maps obtained via IQT-RF, trained with 1.25 mm images downsampled to 2.5 mm, but using the original 1.25 mm image as input for one and two iterations, respectively. The figure fo-

cusses on regions of interest in increasingly (top to bottom) lateral sagittal slices. The top row shows a slice near the midline of the brain and focusses on a region in the cerebellum. IQT-RF reveals the fine tree-like structure of the cerebellar white matter, e.g. the thin pathways ending at the white horizontal arrows, which are difficult to discern at the original resolution. The middle row illustrates clearer depiction after IQT-RF super-resolution of the fine interdigitating pathways in the brain stem at the level of the pons (blue and red area). IQT-RF infers sub-voxel crossing structures (e.g. white vertical upward arrow) and distinct subareas (white vertical downward arrows) obscured by partial volume effects at the original resolution. The bottom row shows that IQT-RF highlights the fine separating pathways in the lateral areas of the corpus callosum (white diagonal arrows) projecting to different gyri; the pattern is consistent with those in very high resolution ex-vivo DTI and polarised light imaging, e.g. [50]. Figure S4 shows equivalent images obtained from cubic interpolation of each DWI prior to fitting the DT. The interpolated images do reveal some of the same hidden detail, but are more blurred. For example, the fine pathways of the cerebellum are less clear than from IQT-RF. Although these results are compelling, validation is challenging due to lack of ground truth. Future validation work with very high resolution data is necessary to confirm efficacy.

IQT-enhanced tractography reveals thin pathways in standard resolution data previously identifiable only in specialised high-resolution data. We concentrate here on just one well-known tractography task to illustrate the potential and further work will evaluate the benefits more generally. However, exploitation of IQT in this way promises major benefits in neuroanatomy [51, 52] and patient connectivity studies [53, 54], where tractography has provided fundamental advances in understanding. It offers similar benefits in neurosurgical planning e.g. [45], which relies increasingly on tractography, but with low-resolution data from lim-

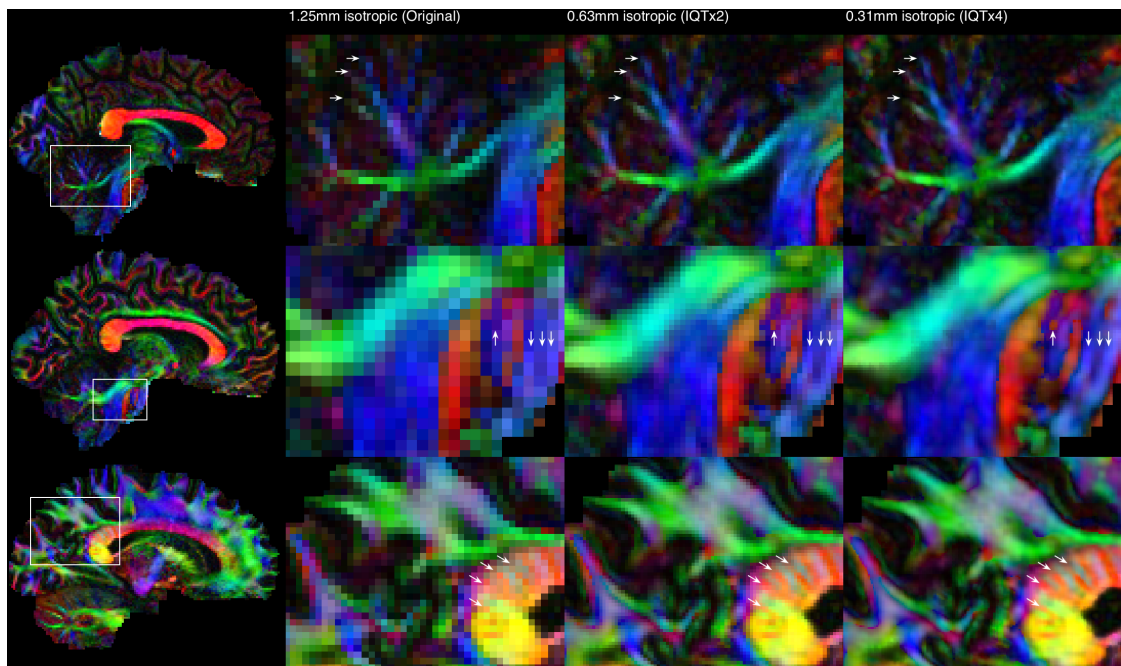


Figure 12: “Super-resolution” DTI maps obtained from IQT-RF. The first column shows DEC maps from three sagittal slices at increasingly lateral positions (top to bottom) from the same data set used for figure 2 at the original 1.25 mm-isotropic resolution. The second column shows the regions of interest marked in the first column at the original 1.25 mm-isotropic resolution. The third and fourth columns show the same regions of interest in 0.625 mm and 0.313 mm isotropic resolution data sets obtained from IQT-RF with $m = 2$ and $n = 2$ after one and two iterations, respectively.

ited acquisition time; Figure 9 shows feasibility of IQT in such scenarios. Moreover, the benefits that IQT offers tractography are mostly complementary to other advances in tractography, such as global and microstructure-informed tractography [55, 56, 57] or machine-learning based tractography [58]. A recent evaluation of signal models in [40] suggests that other signal representations can recover fibre orientations more accurately than MAP-MRI, so future work may examine the impact of the precise signal representation IQT uses and make choices that further enhance performance. **Figure 8 shows that occasionally (e.g. subject 7) tractography matches ground truth better after cubic interpolation than IQT-RF. This occurs despite the IQT-RF reconstructed MAP-MRI map always having lower error compared to ground truth than interpolation. In this particular case, IQT in the seed region relies on boundary completion, which reduces accuracy disrupting the tractography process from the start.** R2.3

IQT parameter mapping gives tantalising results in recovering multi-shell microstructure-imaging parameter maps from single-shell input. The ability to estimate subtle microstructure parameters from single-shell DTI data potentially enables reanalysis of a wide range of historical brain-imaging studies to reveal more specific information on group differences or disease mechanisms. However, this estimation problem is extremely challenging and ill-posed, as the effects of neurite density and orientation dispersion are inseparable in isolated voxels from single-shell input data [18]. IQT manages to resolve much of the ambiguity by exploiting the local image (i.e. patch) structure. For example, within-voxel orientation dispersion from NODDI or SMT must have some consistency with the distribution of orientations in surrounding voxels; IQT-RF with non-trivial patch size learns such constraints implicitly. Nevertheless, the current implementation is not sufficient to replace the multi-shell acquisition. The ellipses in figure 4 highlight a loss of detail. Figure 13 shows how this manifests in a group study: IQT-NODDI recovers

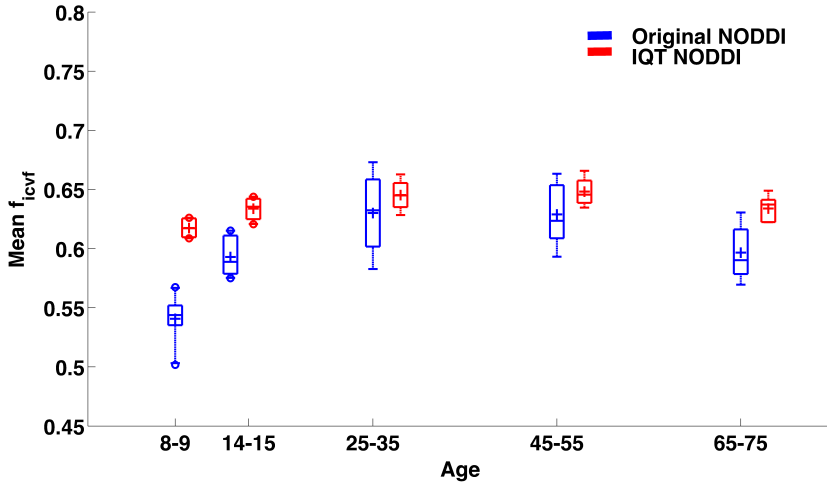


Figure 13: Box plots showing the trends across the different age groups in the Lifespan data set of the NODDI f_{ICVF} parameter from standard NODDI fitting and IQT-RF estimation from DTI after training on HCP1 (8 training images, $|T| \approx 1.6 \times 10^6$, 8 trees, $n = 1$, $m = 1$). The markers show the mean, median, standard deviation, and range over the white matter skeleton from the popular tract-based spatial-statistics method [60].

the known trend across age-groups of NODDI’s f_{ICVF} parameter, which increases to middle age then decreases towards old age [59]. However, the absolute values are biased upwards and the ranges are narrower. This reflects the fact that even considering neighbourhood configurations, the mapping from DTI to NODDI is underdetermined so IQT-RF cannot capture subtle variations in parameter values. Despite these limitations, the current implementation may be sufficient to provide pilot data motivating reacquisition of existing single shell data sets using a multi-shell protocol. Alternative input parameters, such as crossing fibre orientations, T1 or T2-weighted intensity patterns, white-matter-grey-matter classification, or even the weak parameter estimates from the second column of figure 4, may improve performance. In general though, IQT parameter mapping may be more directly useful in stabilising estimates from very sparse, but multi-shell, data sets, as [23] and figure 9 explore. Future work will test the potential further.

Although the simplicity of the current IQT implementations is appealing, many other implementations are feasible and may improve performance. More precise simulation of low-quality from high-quality data to generate matched training pairs is certainly possible [13]. Other learning algorithms, e.g. deep learning via convolutional neural networks [61], may capture more detailed and precise IQT mappings **and may help specifically in learning mappings at the boundary to help tractography from cortical regions**. Enforcing compatibility of neighbouring patches, e.g. via a Markov random field or AutoContext [62], may help resolve remaining ambiguities. For example, close inspection of the 0.313 mm images in figure 12 reveals subtle artefacts, e.g. a blockiness to the image in the vicinity of the upward vertical arrow in the middle row, that arise when pushing the technique to its limit. Neighbourhood constraints and uncertainty evaluation can add further constraints to avoid such effects. **R2.3**

All these candidate IQT implementations support many other important image reconstruction and analysis challenges, such as: synthesising other image contrasts, e.g. estimating T2-weighted images from T1 [13, 14], to reduce acquisition time in clinical studies, or estimating X-ray CT images from MRI [15], to avoid irradiating patients in cancer-therapy planning; learning mappings among imaging protocols to reduce confounds in multi-centre studies or studies that straddle scanner upgrades [10]; or removing artefacts e.g. from signal drop-out or subject motion. The demonstrations we make here illustrate the potential of IQT on its own to bring the power of tomorrow’s imaging techniques into today’s clinical applications. However, even greater future potential lies in IQT’s complementarity to rapid image acquisition strategies, such as compressed sensing [11], MR fingerprinting [12], or simultaneous multislice [30, 63]. The combination offers great promise in realising practical low-power MR, or other imaging, devices such as portable desktop, ambulance or battlefield MRI scanners [64, 65], or in intra-

operative imaging applications with a very tight acquisition-time budget, e.g. [45]. Longer-term, these ideas support a future medical-imaging paradigm exploiting coupled design of a) bespoke high-powered devices to capture databases of high quality images, and b) widely deployed cheap and/or low-power devices designed specifically to exploit the rich information from (a).

Acknowledgements

Microsoft Research and EPSRC grants E007748, I027084, L022680, L023067, M020533 and N018702 supported this work. Data were provided in part by the Human Connectome Project WU-Minn Consortium (Principal Investigators: David Van Essen and Kamil Ugurbil; 1U54MH091657), funded by the 16 NIH Institutes and Centers that support the NIH Blueprint for Neuroscience Research; and by the McDonnell Center for Systems Neuroscience at Washington University. The authors thank Sam Hurley (FMRIB) for his assistance in acquisition of the Prisma data set. We also thank the Behavioral Science Foundation, St. Kitts, for providing the fixated primate brains for the study.

References

- [1] K. Ugurbil, J. Xu, E.J. Auerbach, S. Moeller, A. Vu, J.M. Duarte-Carvajalino, C. Lenglet, X. Wu, S. Schmitter, P. Van de Moortele, J. Strupp, G. Sapiro, F. De Martino, D. Wang, N. Harel, M. Garwood, L. Chen, D.A. Feinberg, S.M. Smith, K.L. Miller, S.N. Sotiropoulos, S. Jbabdi, J.L. Andersson, T.E. Behrens, M.F. Glasser, D. Van Essen, E. Yacoub, and for the WU-Minn HCP Consortium. Pushing spatial and temporal resolution for

- functional and diffusion MRI in the human connectome project. *Neuroimage*, 80:80–104, 2013.
- [2] K. Setsompop, R. Kimmlingen, E. Eberlein, T. Witzel, J. Cohen-Adad, J.A. McNab, B. Keil, M.D. Tisdall, P. Hoecht, P. Dietz, S.F. Cauley, V. Tountcheva, V. Matschl, V.H. Lenz, K. Heberlein, A. Potthast, H. Thein, J. Van Horn, A. Toga, F. Schmitt, D. Lehne, B.R. Rosen, V.J. Wedeen, and L.L. Wald. Pushing the limits of in vivo diffusion MRI for the human connectome project. *NeuroImage*, 80:220 – 233, 2013.
- [3] S. N. Sotiropoulos, S. Jbabdi, J. Xu, J. L. Andersson, S. Moeller, E. J. Auerbach, M. F. Glasser, M. Hernandez, G. Sapiro, M. Jenkinson, D. A. Feinberg, E. Yacoub, C. Lenglet, D. C. Van Essen, K. Ugurbil, and T. E. J. Behrens. Advances in diffusion MRI acquisition and processing in the human connectome project. *NeuroImage*, 80(0):125 – 143, 2013. Mapping the Connectome.
- [4] A.T. Vu, E. Auerbach, C. Lenglet, S Moeller, S.N. Sotiropoulos, S. Jbabdi, J. Andersson, E Yacoub, and K. Ugurbil. High resolution whole brain diffusion imaging at 7T for the human connectome project. *Neuroimage*, 2015.
- [5] H. Greenspan, G. Oz, N. Kiryati, and S. Peled. MRI inter-slice reconstruction using super-resolution. *Magnetic Resonance Imaging*, 20:437–446, 2002.
- [6] P. Coupé, J. V. Manjón, M. Chamberland, M. Descoteaux, and B. Hiba. Collaborative patch-based super-resolution for diffusion-weighted images. *NeuroImage*, 83(0):245 – 261, 2013.
- [7] B. Scherrer, A. Gholipour, and S.K. Warfield. Super-resolution reconstruction to increase the spatial resolution of diffusion weighted images from or-

- thogonal anisotropic acquisitions. *Medical Image Analysis*, 16:1465–1476, 2012.
- [8] G. Van Steenkiste, B. Jeurissen, J. Veraart, A. J. den Dekker, P. M. Parizel, D. H. J. Poot, and J. Sijbers. Super-resolution reconstruction of diffusion parameters from diffusion-weighted images with different slice orientations. *Magnetic Resonance in Medicine*, 75:181–195, 2016.
- [9] L. Ning, K. Setsompop, O. Mikhailovich, N. Makris, M. E. Shenton, C-F Westin, and Y. Rathi. A joint compressed-sensing and super-resolution approach for very high resolution diffusion imaging. *Neuroimage*, 125:386–400, 2016.
- [10] H. Mirzaalian, L. Ning, P. Savadjiev, O. Pasternak, S. Bouix, O. Michailovich, G. Grant, C. E. Marx, R. A. Morey, L. A. Flashman, M. S. George, T. W. McAllister, N. Andaluz, L. Shutter, R. Coimbra, R. D. Zafonte, M. J. Coleman, M. Kubicki, C-F. Westin, M. B. Stein, M. E. Shenton, and Y. Rathi. Inter-site and inter-scanner diffusion mri data harmonization. *Neuroimage*, 135:311–323, 2016.
- [11] M. Lustig, D.L. Donoho, and J.M. Pauly. SparseMRI: the application of compressed sensing for rapid MR imaging. *Magnetic resonance in medicine*, 58:1182–1195, 2007.
- [12] D. Ma, V. Gulani, N. Seiberlich, K. Liu, J. L. Sunshine, J. L. Duerk, and M. A. Griswold. Magnetic resonance fingerprinting. *Nature*, 495:187–192, 2013.
- [13] A. Jog, A. Carass, S. Roy, D.L. Pham, and J.L. Prince. MR image synthesis by contrast learning on neighborhood ensembles. *Medical Image Analysis*, 24:63–76, 2015.

- [14] D. H. Ye, D. Zikic, B. Glocker, A. Criminisi, and E. Konukoglu. Modality propagation: coherent synthesis of subject-specific scans with data-driven regularisation. In K. Mori, I. Sakuma, Y. Sato, C. Barillot, and C. Navab, editors, *MICCAI*, volume 8151 of *LNCS*, pages 606–613, Tokyo, 2013. Springer.
- [15] N. Burgos, M.J. Cardoso, F. Guerreiro, C. Veiga, M. Modat, J. McClelland, A-C Knopf, S. Punwani, D. Atkinson, S.R. Arridge, B.F. Hutton, and S. Ourselin. Robust CT synthesis for radiotherapy planning: Application to the head and neck region. In *Proc. MICCAI*, 2015.
- [16] K. Bahrami, F. Shi, X. Zong, H. W. Shin, H. An, and D. Shen. Reconstruction of 7T-like images from 3T MRI. *IEEE Trans. Med. Im.*, 35:2085–2097, 2016.
- [17] T. B. Dyrby, H. Lundell, M.W. Burke, N.L. Reislev, O.B. Paulson, M. Ptito, and H.R. Seibner. Interpolation of diffusion weighted data sets. *Neuroimage*, 103:202–213, 2014.
- [18] H. Zhang, T. Schneider, C. A. M. Wheeler-Kingshott, and D. C. Alexander. NODDI: Practical in vivo neurite orientation dispersion and density imaging of the human brain. *NeuroImage*, 61:1000–1016, 2012.
- [19] E. Kaden, F. Kruggel, and D. C. Alexander. Quantitative mapping of the per-axon diffusion coefficients in brain white matter. *Magnetic Resonance in Medicine*, 75:1752–1763, 2016.
- [20] E. Kaden, N. D. Kelm, R. P. Carson, M. D. Does, and D. C. Alexander. Multi-compartment microscopic diffusion imaging. *Neuroimage*, 139:346–359, 2016.

- [21] Y. Assaf, D. C. Alexander, Jones. D. K., A. Bizzi, T. E. Behrens, C. A. Clark, Y. Cohen, T. B. Dyrby, P. S. Huppi, T. R. Knösche, D. Le Bihan, G. J. Parker, C. Poupon, CONNECT consortium, D. Anaby, A. Anwender, L. Bar, D. Barazany, T. Blumenfeld-Katzir, S. De-Santis, D. Duclap, M. Figini, E. Fischi, P. Guevara, P. Hubbard, S. Hofstetter, S. Jbabdi, N. Kunz, F. Lazeyras, A. Lebois, M. G. Liptrot, H. Lundell, J. F. Mangin, D. M. Dominguez, D. Morozov, J. Schreiber, K. Seunarine, S. Nava, C. Poupon, T. Riffert, E. Sasson, B. Schmitt, N. Shemesh, S. N. Sotiropoulos, I. Tavor, H. G. Zhang, and F. L. Zhou. The CONNECT project: Combining macro- and micro-structure. *Neuroimage*, 80:273–282, 2013.
- [22] P. J. Basser, J. Mattiello, and D. Le Bihan. MR diffusion tensor spectroscopy and imaging. *Biophysical Journal*, 66:259–267, 1994.
- [23] V. Golkov, A. Dosovitskiy, J. I. Sperl, M. I. Menzel, M. Czisch, P. Sämann, T. Brox, and D. Cremers. q-space deep learning: Twelve-fold shorter and model-free diffusion mri scans. *IEEE Trans. Med. Im.*, 35:1344–1351, 2016.
- [24] J. H. Jensen, J. A. Helpert, A. Ramani, H Lu, and K. Kaczinsky. Diffusional kurtosis imaging: The quantification of non-Gaussian water diffusion by means of magnetic resonance imaging. *Magnetic Resonance in Medicine*, 53:1432–1440, 2005.
- [25] D.C. Van Essen, S.M. Smith, D.M. Barch, T.E. Behrens, E. Yacoub, K. Ugurbil, and WU-Minn HCP Consortium. The WU-Minn human connectome project: an overview. *Neuroimage*, 80:62–79, 2013.
- [26] MF Glasser, SN Sotiropoulos, JA Wilson, TS Coalson, B Fischl, JL Andersson, J Xu, S Jbabdi, M Webster, JR Polimeni, DC Van Essen, M Jenkinson,

- and WU-Minn HCP Consortium. The minimal preprocessing pipelines for the human connectome project. *NeuroImage*, 80:105–124, 2013.
- [27] S. Klöppel, C. M. Stonnington, C. Chu, B. Draganski, R. I. Scahill, J. D. Rohrer, N. C. Fox, C. R. Jack, J. Ashburner, and R. S. J. Frackowiak. Automatic classification of MR scans in Alzheimer’s disease. *Brain*, 131:681–689, 2008.
- [28] T. B. Dyrby, W. F. C. Baaré, D. C. Alexander, J. Jelsing, E. Garde, and L. V. Sjøgaard. An ex-vivo imaging pipeline for producing high quality and high resolution diffusion weighted imaging datasets. *Human Brain Mapping*, 32:544–563, 2011.
- [29] T. B. Dyrby, L. V. Sjøgaard, M. G. Hall, M. Ptito, and D. C. Alexander. Contrast and stability of the axon diameter index from microstructure imaging with diffusion MRI. *Magnetic Resonance in Medicine*, 70:711–721, 2013.
- [30] S. Moeller, E. Yacoub, C.A. Olman, E. Auerbach, J. Strupp, N. Harel, and K. Ugurbil. Multiband multislice GE-EPI at 7 tesla, with 16-fold acceleration using partial parallel imaging with application to high spatial and temporal whole-brain fMRI. *Magnetic Resonance in Medicine*, 63:1144–1153, 2010.
- [31] E. Özarlan, C.G. Koay, T.M. Shepherd, M.E. Komlosh, M.O. Irfanoglu, C. Pierpaoli, and P. Basser. Mean apparent propagator (MAP) MRI: A novel diffusion imaging method for mapping tissue microstructure. *Neuroimage*, 78:16–32, 2013.
- [32] H. E. D’Arceuil, S. Westmoreland, and A. J. de Crespigny. An approach to high resolution diffusion tensor imaging in fixed primate brain. *Neuroimage*, 35:553–565, 2007.

- [33] D. K. Jones and P. J. Basser. Squashing peanuts and smashing pumpkins: How noise distorts diffusion-weighted MR data. *Magnetic Resonance in Medicine*, 52:979–993, 2004.
- [34] L. Breiman. Random forests. *Machine learning*, 45:5–32, 2001.
- [35] A. Criminisi and J. Shotton. *Decision forests for computer vision and medical image analysis*. Springer, 2013.
- [36] C.-F. Westin, S. E. Maier, B. Khidhir, P. Everett, F. A. Jolesz, and R. Kikinis. Image processing for diffusion tensor magnetic resonance imaging. In *Proc. MICCAI*, pages 441–452, Cambridge, 1999. Springer.
- [37] D. C. Alexander and G. J. Barker. Optimal imaging parameters for fibre-orientation estimation in diffusion MRI. *NeuroImage*, 27:357–367, 2005.
- [38] S.J.D. Prince. *Computer Vision: Models, Learning, and Inference*. Cambridge University Press, 2012.
- [39] S. Jbabdi, S. N. Sotiropoulos, A. M. Savio, M. Grana, and T. E. J. Behrens. Model-based analysis of multishell diffusion MR data for tractography: how to get over fitting problems. *Magnetic Resonance in Medicine*, 68:1846–1855, 2012.
- [40] L. Ning, F. Laun, Y. Gur, E.V.R. DiBella, S. Deslauriers-Gauthier, T. Megherbi, A. Ghosh, M. Zucchelli, G. Menegaz, R. Fick, S. St-Jean, M. Paquette, R. Aranda, M. Descoteaux, R. Deriche, L. O’Donnell, and Y. Rathi. Sparse reconstruction challenge for diffusion mri: Validation on a physical phantom to determine which acquisition scheme and analysis method to use? *Medical Image Analysis*, 26:316–331, 2015.

- [41] S. Pajevic and C. Pierpaoli. Color schemes to represent the orientation of anisotropic tissues from diffusion tensor data: application to white matter fiber tract mapping in the human brain. *Magnetic Resonance in Medicine*, 42:526–540, 1999.
- [42] P. J. Basser and C. Pierpaoli. Microstructural and physiological features of tissues elucidated by quantitative diffusion tensor MRI. *Journal of Magnetic Resonance Series B*, 111:209–19, 1996.
- [43] D. E. Haines. *Neuroanatomy: an atlas of structures, sections, and systems*. Lippincott, Williams and Wilkins, 2012.
- [44] S. J. Counsell, M. C. Allsop, J. M. and Harrison, D. J. Larkman, N. L. Kennea, O. Kapellou, F. M. Cowan, J. V. Hajnal, A. D. Edwards, and M. A. Rutherford. Diffusion-weighted imaging of the brain in preterm infants with focal and diffuse white matter abnormality. *Pediatrics*, 112:1–7, 2003.
- [45] G. P. Winston, P. Daga, J. Stretton, M. Modat, M. R. Symms, A. W. McEvoy, S. Ourselin, and J. S. Duncan. Optic radiation tractography and vision in anterior temporal lobe resection. *Annals of Neurology*, 71:334–341, 2012.
- [46] J. Anderson and S.N. Sotiropoulos. An integrated approach to correction for off-resonance effects and subject movement in diffusion mr imaging. *Neuroimage*, 125:1063–1078, 2016.
- [47] J.V. Manjon, P. Coupe, A. Buades, V. Fonov, D. Louis Collins, and M. Robles. Non-local mri upsampling. *Medical Image Analysis*, 14:784–792, 2010.
- [48] R. Tanno, A. Ghosh, F. Grussu, E. Kaden, A. Criminisi, and D. C. Alexander. Bayesian image quality transfer. In *Proc. MICCAI*, pages 265–273, 2016.

- [49] John Blitzer, Ryan McDonald, and Fernando Pereira. Domain adaptation with structural correspondence learning. In *Proceedings of the 2006 conference on empirical methods in natural language processing*, pages 120–128. Association for Computational Linguistics, 2006.
- [50] J. Mollink, M. Kleinnijenhuis, S. N. Sotiropoulos, M. Cottaar, A.-M. van Cappellen van Walsum, M. Gamarallage, O. Ansorge, S. Jbabdi, and K. Miller. Exploring fibre orientation dispersion in the corpus callosum: Comparison of dMRI, PLI and histology. In *Proc. OHBM*, page 2056. OHBM, 2016.
- [51] M. Catani, D. K. Jones, and D. H. ffytche. Perisylvian language networks of the human brain. *Annals of Neurology*, 57:8–16, 2005.
- [52] V. J. Wedeen, D. L. Rosene, R. Wang, G. Dai, F. Mortazavi, P. Hagmann, J. H. Kaas, and W.-Y. I. Tseng. The geometric structure of the brain fiber pathways. *Science*, 335:1628–1634, 2012.
- [53] M. Catani and D. H. ffytche. The rises and falls of disconnection syndromes. *Brain*, 128:2224–2239, 2005.
- [54] J. D. Warren, J. D. Rohrer, J. M. Schott, N. C. Fox, J. Hardy, and M. N. Rossor. Molecular nexopathies: a new paradigm of neurodegenerative disease. *Trends in neurosciences*, 36:561–569, 2013.
- [55] A. J. Sherbondy, M. C. Rowe, and D. C. Alexander. MicroTrack: an algorithm for concurrent projectome and microstructure estimation. In *Proc. MICCAI, LNCS 6361*, pages 183–190, Beijing, 2010. Springer.
- [56] M. Reisert, V. G. Kiselev, B. Dihtal, E. Kellner, and D. S. Novikov. MesoFT: unifying diffusion modelling and fiber tracking. In *MICCAI*, pages 201–208, 2014.

- [57] A. Daducci, A. Dal Palu, A. Lemkadden, and J.P. Thiran. COMMIT: Convex optimization modeling for microstructure informed tractography. *IEEE Trans. Med. Im.*, 34:246–257, 2015.
- [58] P. F. Neher, M. Götz, T. Norajitra, C. Weber, and K. H. Maier-Hein. A machine learning based approach to fiber tractography using classifier voting. In *Proc. MICCAI*, 2015.
- [59] Y. S. Chang, J. P. Owen, N. J. Pojman, T. Thieu, P. Bukshpun, M. L. J. Wakahiro, J. I. Berman, T. P. L. Roberts, S. S. Nagarajan, E. H. Sherr, and P. Mukherjee. White matter changes of neurite density and fiber orientation dispersion during human brain maturation. *PLoS ONE*, 10(6):e0123656, 2015.
- [60] S. M. Smith, M. Jenkinson, H. Johansen-Berg, D. Rueckert, T. E. Nichols, C. E. Mackay, K. E. Watkins, O. Ciccarelli, M. Zaheer Cadera, P. M. Matthews, and T. E. J. Behrens. Tract-based spatial statistics: Voxelwise analysis of multi-subject diffusion data. *Neuroimage*, 31:1487–1505, 2006.
- [61] C. C. Dong, C. Loy, K. He, and X. Tang. Image super-resolution using deep convolutional networks. *IEEE Trans. Pattern Analysis and Machine Intelligence*, 38:295–307, 2016.
- [62] Z. Tu and X. Bai. Auto-context and its application to high-level vision tasks and 3d brain image segmentation. *IEEE Trans. Pattern Anal. Mach. Intell.*, 32:1744–1757, 2010.
- [63] K. Setsompop, B.A. Gagoski, J.R. Polimeni, T. Witzel, V.J. Wedeen, and L.L. Wald. Blipped-controlled aliasing in parallel imaging for simultaneous multislice echo planar imaging with reduced g-factor penalty. *Magnetic resonance in medicine*, 67:1210 – 1224, 2012.

- [64] C. Z. Cooley, J. P. Stockmann, B. D. Armstrong, M. Sarracanie, M. H. Lev, M. S. Rosen, and L. L. Wald. Two-dimensional imaging in a lightweight portable MRI scanner without gradient coils. *Magnetic Resonance in Medicine*, 73:872–883, 2015.
- [65] M. Sarracanie, C. D. LaPierre, N. Salameh, D. E. J. Waddington, T. Witzel, and M. S. Rosen. Low-cost high-performance MRI. *Scientific Reports*, 5:15177, 2015.

Supplementary material

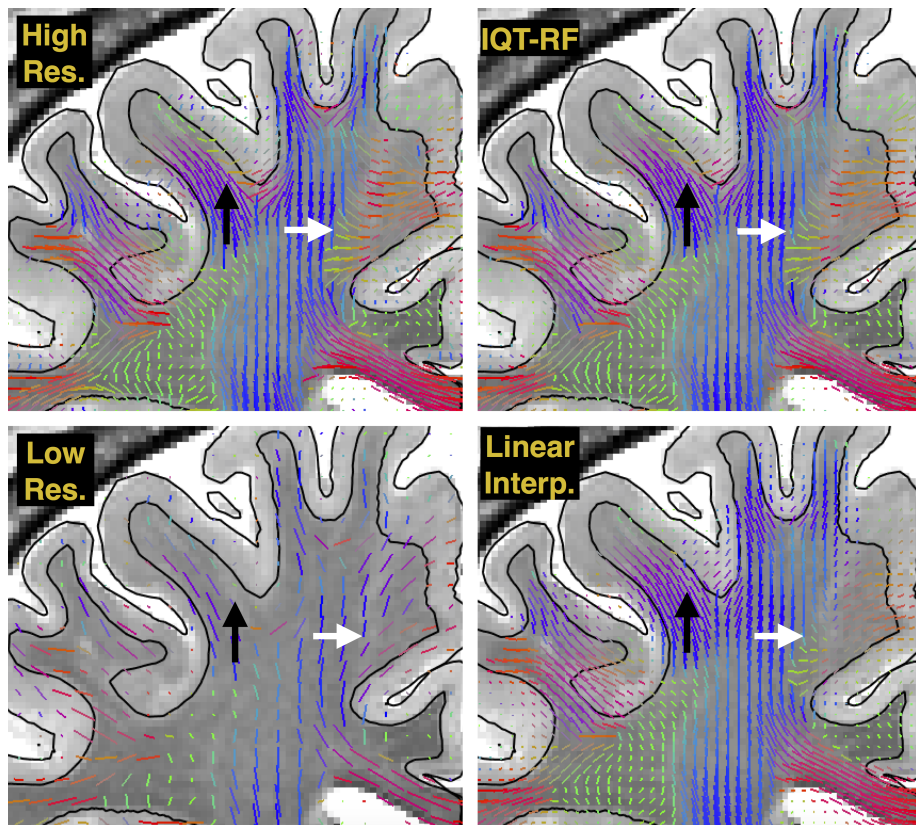


Figure S1: As figure 3, but without boundary completion, i.e. voxels that correspond to incomplete low-res input patches have no content. Comparison with figure 3 shows where boundary completion comes into play and enables visual assessment of its efficacy. For example, in this figure, vectors are missing at the crown of the gyrus containing the black arrow. At that crown figure 3 shows that the visual match of IQT-RF output to ground truth is still closer than that from interpolation.

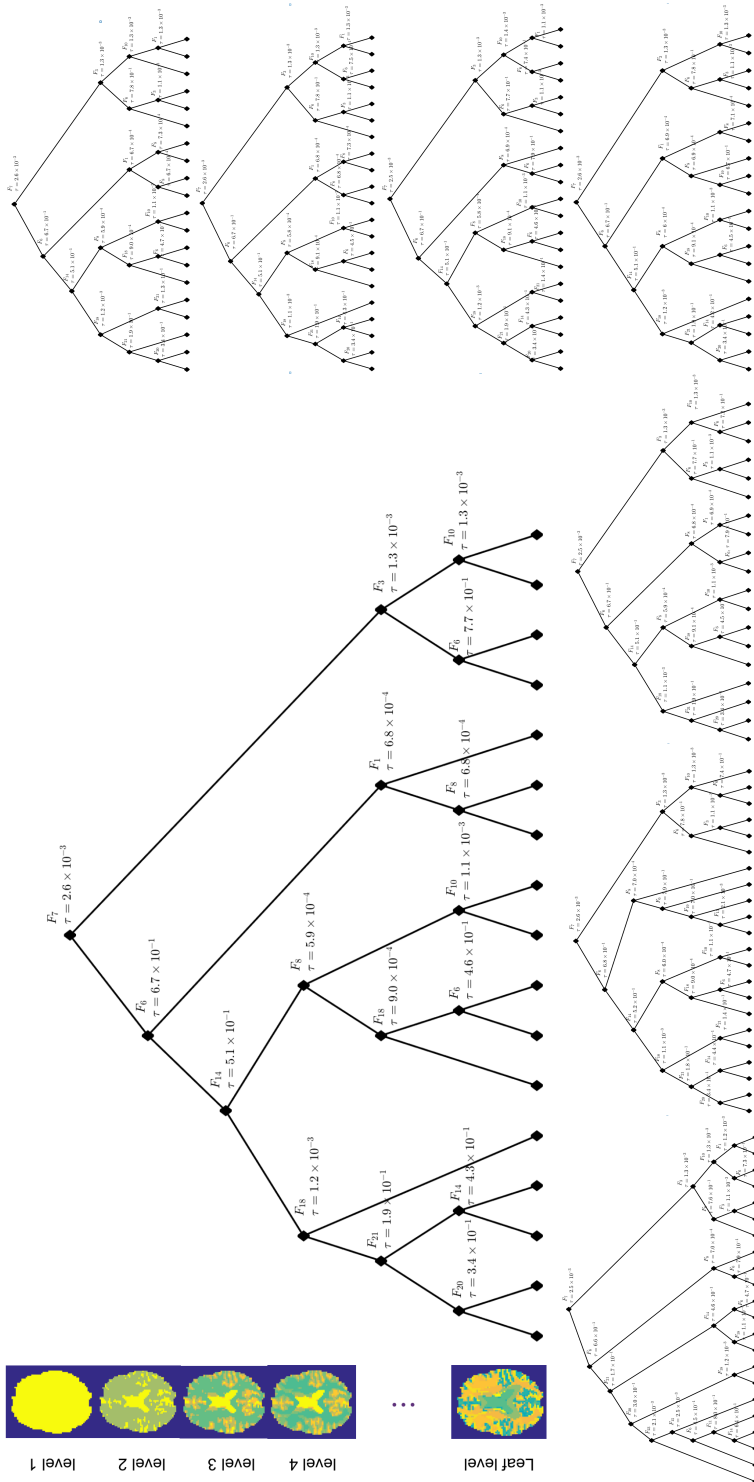


Figure S2: Sketch of the eight trees comprising the forest trained for resolution enhancement on HCP1 with $|T| \approx 1.6 \times 10^6$, $m = 2$, and $n = 2$. The maps to the left of the big tree show, in an example slice, how different tissue types are assigned to different branches. The label at each node gives the feature and threshold that determine the decision at that node. F_1 to F_3 are the eigenvalues (from largest to smallest) of the DT at the patch centre; F_8 to F_{10} are the means of each eigenvalue over the central $3 \times 3 \times 3$ patch; and F_{16} to F_{18} are the means over the whole patch. F_4 to F_7 are the linearity, planarity, isotropy, and trace of the central DT; F_{12} to F_{15} are the means of those features over the central $3 \times 3 \times 3$ patch; and F_{20} to F_{23} are the means over the whole patch. Finally, F_{11} and F_{19} are measures of orientation dispersion over the central $3 \times 3 \times 3$ and whole patch, respectively. The units of eigenvalue and trace features are $\text{mm}^2 \text{s}^{-1}$.

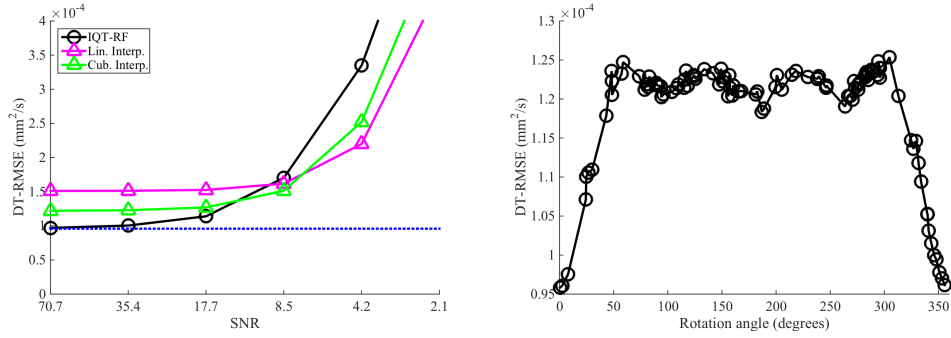


Figure S3: The left panel plots DT-RMSE (for HCP1 training and testing, all 8 subjects in both training and test sets, $m = 2$, $n = 2$, $|T| \approx 1.6 \times 10^6$) against SNR of the input data. The right panel plots DT-RMSE (similarly) against simulated head-rotation angle.

For IQT, each data point on the left panel comes from adding noise to the downsampled DWIs, fitting the DT in each voxel, and reconstructing the high resolution DTI through IQT-RF (black markers). Interpolation (green and magenta markers) is performed on the DWIs prior to DT estimation. The noise is Rician distributed with level yielding the SNR specified on the x-axis on the mean unweighted intensity in the brain parenchyma. The dotted blue line shows the DT-RMSE for IQT-RF without additional noise. IQT-RF performance degrades gracefully with the addition of noise. It continues to outperform both linear and cubic interpolation until the SNR drops below about 9 and about 12, respectively. Note, however, that the uncorrupted image is not noise free so noise levels are a little higher than indicated. For noisy input data, including artificial noise corruption in the training may improve robustness; future work will test this assertion.

Each of the 100 data points on the right panel comes from rotating all DTs in the input data set by a random rotation matrix (the same one for the whole image), reconstructing the high resolution images through IQT-RF, and subsequently rotating by the inverse rotation. This reflects the ability of the IQT-RF mapping to operate on an image acquired with the head orientation corresponding to the random rotation. The plot shows that performance drops as the orientation departs from the fixed orientation of the HCP data sets and stabilises at about 45° with performance around the level of cubic interpolation. In applications in which head orientation varies substantially, inclusion of multiple orientations in the training set would be important to make the mapping more robust.

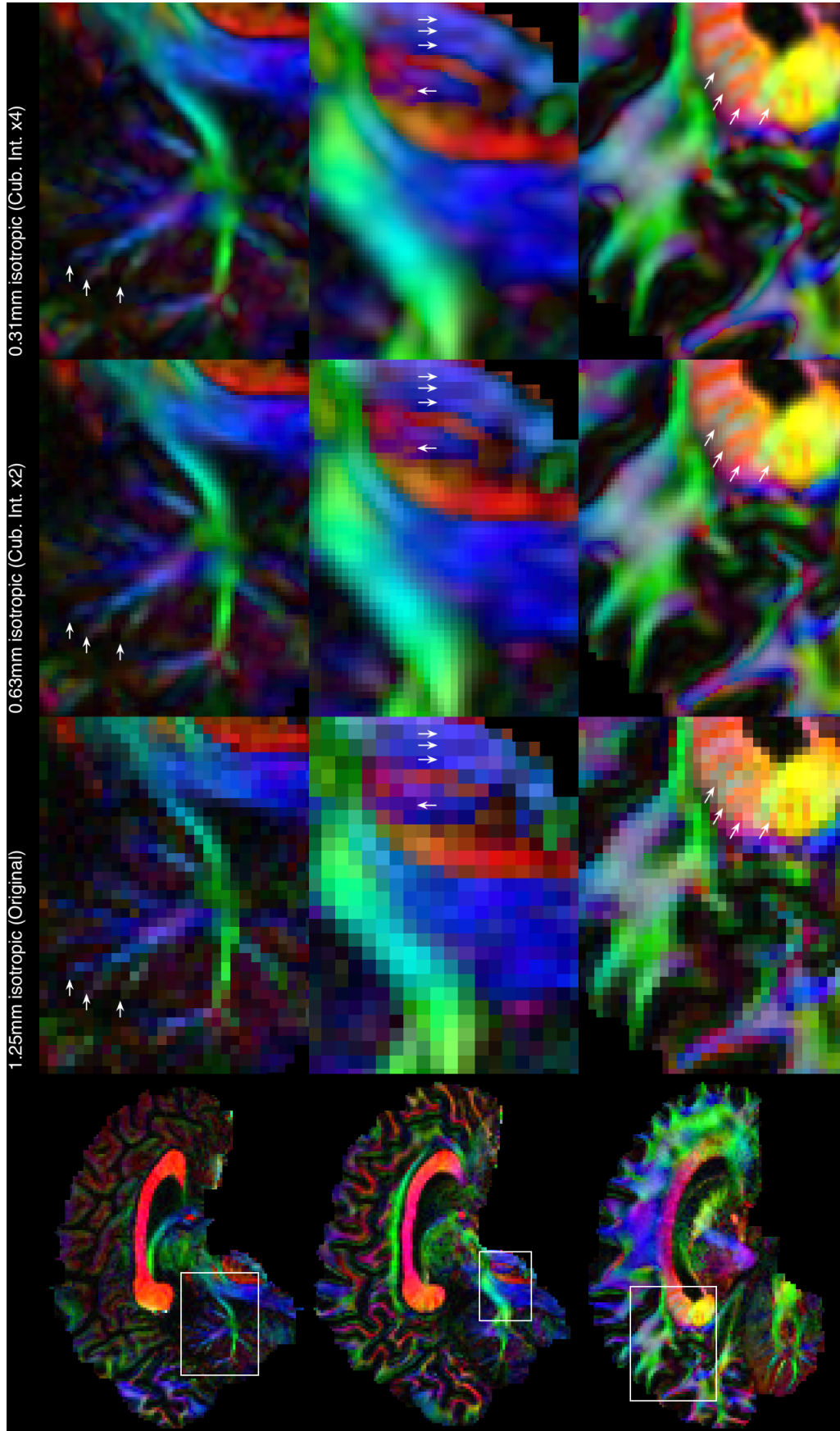


Figure S4: Super-resolution DTIs obtained from cubic interpolation of each component DWI for comparison with the IQT super-resolution maps in figure 12. The interpolated images show less sharp detail than those obtained from IQT.