Image Quality Transfer via Random Forest Regression: Applications in Diffusion MRI

Daniel C. Alexander[1], Darko Zikic[2], Jiaying Zhang[1], Hui Zhang[1], and Antonio Criminisi[2]

[1] Centre for Medical Image Computing and Department of Computer Science, University College London, Gower Street, London, WC1E 6BT, UK
[2] Microsoft Research Cambridge, Cambridge, UK

Abstract. This paper introduces image quality transfer. The aim is to learn the fine structural detail of medical images from high quality data sets acquired with long acquisition times or from bespoke devices and transfer that information to enhance lower quality data sets from standard acquisitions. We propose a framework for solving this problem using random forest regression to relate patches in the low-quality data set to voxel values in the high quality data set. Two examples in diffusion MRI demonstrate the idea. In both cases, we learn from the Human Connectome Project (HCP) data set, which uses an hour of acquisition time per subject, just for diffusion imaging, using custom built scanner hardware and rapid imaging techniques. The first example, super-resolution of diffusion tensor images (DTIs), enhances spatial resolution of standard data sets with information from the high-resolution HCP data. The second, parameter mapping, constructs neurite orientation density and dispersion imaging (NODDI) parameter maps, which usually require specialist data sets with two b-values, from standard single-shell high angular resolution diffusion imaging (HARDI) data sets with $b = 1000\,\mathrm{s\,mm^{-2}}$. Experiments quantify the improvement against alternative image reconstructions in comparison to ground truth from the HCP data set in both examples and demonstrate efficacy on a standard data set.

1 Introduction

Bespoke MRI scanners and imaging protocols can produce very high quality data uniquely informative about anatomy and function. However, the imaging techniques that underpin such data sets are often impossible or impractical on standard devices or in clinical imaging scenarios. For example, small-animal scanners often have higher field and gradient strength and smaller bore than human scanners, enhancing signal to noise, image resolution and, in diffusion MRI, sensitivity to small structures. Such platforms can provide exquisitely high resolution images revealing fine structural detail and providing strong sensitivity to anatomical features or pathology. Although such measurements highlight the potential of future human imaging devices, they provide little direct benefit to current clinic practice. Similarly, the HCP designed bespoke MRI scanners equipped

P. Golland et al. (Eds.): MICCAI 2014, Part III, LNCS 8675, pp. 225–232, 2014.

with $100\,\mathrm{mT\,m^{-1}}$ and $300\,\mathrm{mT\,m^{-1}}$ gradient coils (several times more powerful than standard clinical scanners) and exploit several imaging and image reconstruction innovations to speed up acquisition and improve data quality [1]. The bespoke imaging system combined with a lengthy acquisition protocol leads to unique data sets with unprecedented image resolution and noise levels. However, the techniques extend only partially to clinical imaging with modest hardware and much more limited imaging times.

In this paper, we propose to exploit the information in expensive high quality data sets to improve images reconstructed from more modest data acquisitions. We call this process *image quality transfer*. We learn fine image structure from high quality data sets and use it to enhance lower quality data. We present a framework for solving this general problem using a patch-based image representation and random forest regression.

Two distinct applications demonstrate the framework by exploiting the HCP in-vivo human diffusion MRI data, which has uniquely high quality. The first application is super-resolution of DTIs. The HCP diffusion data have voxel sizes of $1.25^3\,\mathrm{mm^3}$ rather than typical sizes around $2^3\,\mathrm{mm^3}$ in standard data sets. Image quality transfer provides a mechanism to reconstruct high resolution DTIs from low-resolution acquisitions. The second application is quantitative parameter mapping. The HCP provides three HARDI shells of data with diffusion weighting factor $b \approx 1000$, 2000, and $3000\,\mathrm{s\,mm^{-2}}$, which supports estimation of more informative parameters than standard data sets with a single HARDI shell at $b = 1000\,\mathrm{s\,mm^{-2}}$. For example, NODDI [2] provides more specific information than DTI, such as maps of the density and dispersion of neurites (axons and dendrites), by fitting a more informative model in each voxel. It has become popular for clinical studies, because it requires as little as 15 minutes acquisition time. However, NODDI requires at least two HARDI shells with distinct b and fitting the NODDI model fails with only a single HARDI shell [2]. This prevents its use on the large variety of historical standard data sets. Image quality transfer provides a mechanism to recover NODDI parameter maps from single b-value data sets, which potentially enables NODDI analysis of historical data.

Prior literature on super-resolution is extensive. In medical imaging, [3] uses example patches from high resolution images to super-resolve scalar MR images and [4] use dictionaries from a database of similar images. Several authors propose super-resolution techniques specifically for diffusion images. The closest to our work is [5], which enhances the resolution of each diffusion weighted image (DWI) through patch examples before fitting the DT or other models; the discussion compares this approach to ours in more detail. Image quality transfer for parameter mapping from rarefied data sets holds greater novelty, although similar in spirit to modality transfer [6], which predicts T2 and FA images from T1 scans via patch-based label propagation. Our framework solves both problems.

2 Methods

Our implementation of image quality transfer learns a mapping from each neighbourhood of N_1 voxels in the low-quality data set to a corresponding

neighbourhood of N_2 voxels in the high-quality data set. Input and output voxels are vector-valued containing p_1 and p_2 values, respectively. Construction of the mapping is thus a regression problem. It requires a training set of patch pairs $T = \{\mathbf{x}_i, \mathbf{y}_i\}_{i=1}^{|T|}$, where each $\mathbf{x}_i$ has dimension $p_1 N_1$ and $\mathbf{y}_i$ dimension $p_2 N_2$.

2.1 Regression Models

We consider a hierarchy of three types of mapping, where each generalises the previous: global linear; regression trees; and regression forests.

For global linear regression, we compute the linear transformation matrix $G = YX^\dagger$, where Y has columns $\mathbf{y}_i$, X has columns $\mathbf{x}_i$, and $X^\dagger$ is an appropriate pseudo inverse of X so that, in matlab, `G = X\Y`. For input patch $\mathbf{x}$, the estimate of the corresponding output patch is $G\mathbf{x}$.

The regression tree implements a piecewise linear regression over the space of input data points [7,8]. Each internal node in the tree sends data points into left or right subtrees by thresholding one of J scalar functions of $\mathbf{x}$, or *features*, $F_1, \cdots, F_J$. The choice of features is application dependent and we define ours later. Each leaf node contains a linear transformation with the same structure as the global linear transformation G defined in the previous section. Thus, for input data point $\mathbf{x}$, the output estimate is $G_t\mathbf{x}$ where G_t is the linear transformation of the leaf node at which the data point arrives after traversing the tree. Training uses a standard greedy search strategy similar to [8]. To control for overfitting, we use a validation set V with similar size to T and accept only splits that reduce the residual error of V.

Regression forests use multiple regression trees constructed from different training sets. Outputs are element by element averages of the prediction from each tree weighted by the error covariance of the linear transformation G_t, estimated during training.

2.2 Application 1: Super-Resolution

For DTI super-resolution, the mapping takes a $(2n+1) \times (2n+1) \times (2n+1)$ cubic patch of DTs, so that $N_1 = (2n+1)^3$ and $p_1 = 6$, as input, and outputs an $m \times m \times m$ cubic patch of voxels, each also containing a DT, so that $N_2 = m^3$ and $p_2 = 6$. The output patch is a cubic array of subvoxels that super-resolves the central voxel of the input patch. For example, if $n = 2$ and $m = 2$, each input $\mathbf{x}$ contains the 6 independent elements of the DT in each voxel of a $5 \times 5 \times 5$ low-resolution patch and each output $\mathbf{y}$ the elements of each DT in the $2 \times 2 \times 2$ high resolution patch (see figure 1); the mapping is thus from $\mathcal{R}^{750}$ to $\mathcal{R}^{48}$.

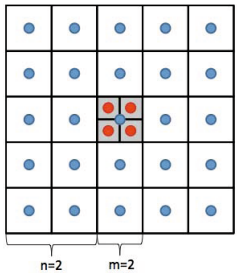

Fig. 1. 2D illustration of the input (blue) and output (red) patch structure for $n = 2$ and $m = 2$

For $F_1, \cdots, F_J$, we use the following features of $\mathbf{x}$:

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

- The orientation of the principal eigenvector of the DT in the central voxel.
- The means of each eigenvalue over the central $3 \times 3 \times 3$ cube and those over the whole $(2n+1)^3$ cube.
- The mean principal orientation over the central 3^3 and the whole cubes.
- The orientational variance over the central 3^3 and the whole cubes.

Unless otherwise stated, training data comes from 8 randomly selected HCP data sets `www.humanconnectome.org`. A separate test set contains a different 8 HCP data sets. Each data set contains 288 DWIs including 18 with $b \approx 0$ and three HARDI shells of 90 directions with $b \approx 1000$, 2000, and $3000\,\mathrm{s\,mm^{-2}}$; the precise values vary spatially, as described in [1]. This application uses only the $b \approx 0$ and $b \approx 1000\,\mathrm{s\,mm^{-2}}$ measurements to reflect standard data sets. Training pairs come from downsampling each DWI by a factor of m in each dimension, fitting the DT in each voxel of both the downsampled and full resolution image using weighted linear least squares accounting for the spatially varying b and gradient directions, 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.

Each data set contains around 7.5×10^5 brain voxels. We randomly subsample the set of patches from the full training set to meet memory limitations.

2.3 Application 2: Parameter Mapping

This application aims to estimate maps of NODDI parameters from standard data sets including only $b \approx 0$ and $b \approx 1000\,\mathrm{s\,mm^{-2}}$ measurements. Thus the mapping takes as input a $(2n+1)^3$ cubic patch of DTs fitted to a $b \approx 1000\,\mathrm{s\,mm^{-2}}$ HARDI shell, so that $N_1 = (2n+1)^3$ and $p_1 = 6$. The mapping outputs the NODDI parameters, intra-cellular volume fraction f_{ICVF}, free-water volume fraction f_{ISO}, orientation dispersion index (ODI), and the mean fibre orientation (θ, ϕ), at the central voxel of the input patch; $N_2 = 1$ and $p_2 = 5$.

The features, $F_1, \cdots, F_J$, training, and test data sets are as in DTI super-resolution. The ground truth NODDI output for the training set comes from fitting the NODDI model to all three HARDI shells in each image voxel.

3 Results

Figure 2 shows qualitative results from both applications on one of the test data sets (not used in training). The left panel compares various super-resolution reconstructions, obtained with $m = 4$, after downsampling by a factor of 4 in each dimension. The right panel of figure 2 compares various reconstructed NODDI maps with ground truth. In both examples, $n = 2$, the training set T contains about 1.5M data points, and the forest uses 8 trees.

For super-resolution, clarity in both the global linear and forest reconstructions compare favourably to standard interpolation techniques. The global linear and forest reconstructions appear quite similar, although the latter avoids some

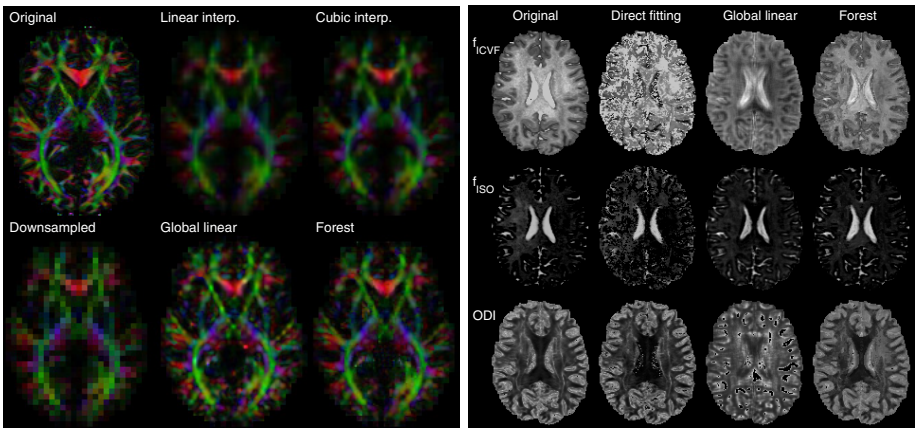

Fig. 2. Left: Direction-encoded colour FA maps for various reconstructed DTIs from a downsampled image (bottom left) compared to ground truth (top left) from the original full resolution data set. Right: Comparison of ground truth NODDI parameter maps (left) with standard fitting to just the $b \approx 1000\,\mathrm{s\,mm^{-2}}$ shell (left middle), global linear (right middle) and random forest (right) regression.

glitches visible in the former. As [2] predicts, the standard voxel-by-voxel NODDI parameter estimation fails, strongly disrupting the neurite density parameter, f_{ICVF}, in particular. The global linear transformation also performs poorly and fails to recover the structure of the f_{ICVF} and ODI maps, although it does produce a reasonable reconstruction of f_{ISO}. The improvement of the forest over the global linear transformation is striking and the output is visually much closer to the ground truth, although some differences are still clearly discernible.

Figure 3 quantifies the comparison of high resolution (left) and NODDI parameter (right) reconstructions as a function of training set size. The metric of reconstruction error is the mean (over the 8 test subjects) median (over brain voxels) root-mean-squared parameter error (i.e. of the six independent DT elements or 3 scalar NODDI parameters). All regression techniques improve on standard interpolation methods for super-resolution. Trees and forests improve on global linear transformations increasingly as the training set size increases, because they exploit additional training data by increasing the complexity of the mapping model. Averaging over the output of multiple trees (the forest) shows benefit over single trees. The advantage comes in part because the forests see more training data than single trees, as each component tree uses a different random training set, although they also mitigate the greedy search that trains individual trees. Available computer memory limits the amount of training data a single tree can use, so the forests offer genuine advantages by enabling exploitation of more training data.

Between subject standard deviation is consistent among the different algorithms suggesting that the error score depends on individual anatomical features,

such as the size of the ventricles where the error scores tend to be largest for all algorithms. Other DT metrics, such as orientational difference, show similar trends. The standard deviation over training set randomisation is more meaningful than over subjects for evaluating the significance of differences in error score. That standard deviation is 3 or 4 orders of magnitude smaller than the error itself in most cases, which suggests that (a) the results are highly reproducible over different training sets; and (b) the differences in error scores between, e.g. trees and forests, are highly significant (around 60 standard deviations).

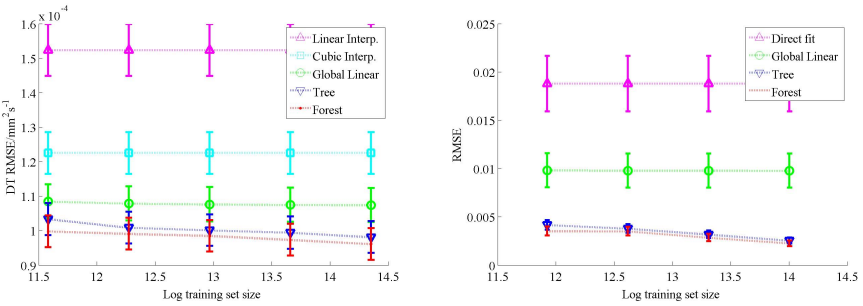

Fig. 3. Left: reconstruction errors against ground truth for various reconstructed DT maps, after downsampling by a factor of 2 and reconstructing with $m = 2$, as a function of training set size. Right: reconstruction errors for NODDI parameter reconstructions. In both cases, $n = 2$, and the error bars show standard deviation over subjects.

Figure 4 shows qualitative results of both image quality transfer applications using mappings learned from HCP data to enhance a non-HCP data set. The data set comes from a standard 3T clinical scanner. It has a single HARDI shell with $b = 1000\,\mathrm{s\,mm^{-2}}$ and 30 gradient directions. The voxel size is $2^3\,\mathrm{mm}^3$. Image quality transfer sharpens weak structures in super-resolution and produces plausible NODDI parameter maps.

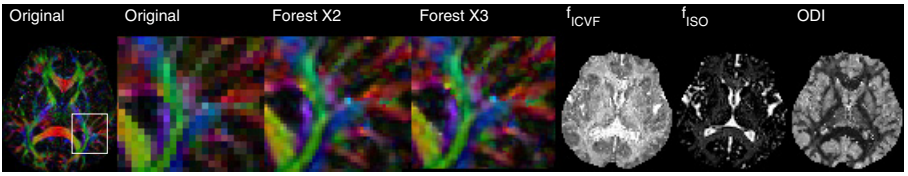

Fig. 4. Image quality transfer to a non-HCP data set. From left to right: (i) Colour FA map of original data; (ii) zoomed in view of boxed area; (iii) and (iv) forest up sampling with $m = 2$ and $m = 3$; (v)-(vii) reconstructed NODDI parameter maps.

4 Discussion

The super-resolution results show that patches in low resolution diffusion images contain a significant amount of information about the subvoxel content at the centre of the patch. Thus, with a transformation of sufficient complexity, we can predict high resolution images with much greater accuracy than standard interpolation techniques. Further work is required to make a formal performance comparison with other super-resolution techniques, in particular [5], but a key novelty of our implementation is to use DT patches as input rather than DWI patches and to output patches of parameter estimates rather than raw DWIs. An advantage of outputting fitted model patches is to constrain the output image structure to realistic local configurations. Moreover, treating each DWI separately, as in [5], fails to exploit the strong correlation among DWIs, which is highly informative. For example, enhancing the raw DWIs and then fitting the NODDI model fails because it still fits to only a single b; it is the neighbourhood structure of the fitted DTs that informs on the parameters. One might consider using patches of the full collection of raw DWIs as input or output. However, this has two practical problems: i) it complicates transfer to sparser data sets where the set of input DWIs is, by definition, different to that of the high quality training data; ii) the memory requirements for training become orders of magnitude larger, as, in the super-resolution example, p_1 increases from 6 to 288. However, input patches of parameters of more complex models than the DT, such as higher-order spherical harmonic coefficients, may improve performance.

Random-forest regression shows a dramatic improvement over both standard parameter estimation and global linear regression in recovering NODDI parameter maps from single-shell HARDI data sets and produces plausible maps of all three important parameters.

Training times can be considerable for regression trees (around 1 day of processor time for the largest training sets in figure 3 with unoptimised matlab code), but reconstruction times are small (a few minutes for a full volume), as they require just a linear transformation in each voxel. This is much quicker in fact than direct NODDI fitting, which is non-linear and requires several hours of processor time per image volume. The method extends naturally to predict parameters of other models, such muti-fibre models [9], where computational advantages again are potentially significant. More broadly, the framework extends naturally beyond diffusion imaging to any scalar or vector valued images.

Imperfections remain in reconstructed images in both applications. Various refinements of the random forest regression may improve performance. Reconstruction error reduces rapidly as the number of trees increases from 1 to 3, but stabilises above 4 so that the 8 trees we use here is sufficient. For fixed training set size, little improvement arises from using more than 2 source images, so our collection of 8 seems sufficient. Performance increases with n for a fixed number of data points in the training set, suggesting that neighbourhoods are informative even several steps from the output voxel. However, the data points are larger at higher n, so memory limits occur at fewer data points; $n = 2$ is a good compromise with the current implementation. Choice of features affects perfor-

mance. From our set, features based on the largest and smallest DT eigenvalues dominate near the tree root; the second eigenvalue rarely appears. Orientational features are important for finer partitions nearer leaf nodes. Smaller n leads to more complex trees, because each linear transformation has less parameters. Other high dimensional regression techniques may improve results. Moreover, constraining the mapping with a data fitting term would be beneficial, but leads to non-linear reconstruction increasing computation times.

In summary, this initial formulation and demonstration of image quality transfer shows compelling results from a simple implementation that improves significantly on standard interpolation and estimation techniques. Further work must establish sufficiency of reliability to make downstream improvements in practical applications such as tractography and image-based biomarker studies. Both random-forest regression and the local patch-based image representation lend themselves well to generalizability in the presence of pathology or other effects not represented in the training set at least for diffuse or macroscopic effects. However, further work needs to evaluate performance in such situations.

Acknowledgements. EPSRC grants E007748 and I027084 supported this work. Data were provided [in part] by the HCP, WU-Minn Consortium (PIs: David Van Essen and Kamil Ugurbil; 1U54MH091657) funded by NIH and Wash. U.

References

1. Sotiropoulos, S.N., et al.: Advances in diffusion MRI acquisition and processing in the human connectome project. NeuroImage 80, 125–143 (2013)
2. Zhang, H., et al.: NODDI: Practical in vivo neurite orientation dispersion and density imaging of the human brain. NeuroImage 61, 1000–1016 (2012)
3. Rousseau, F.: Brain Hallucination. In: Forsyth, D., Torr, P., Zisserman, A. (eds.) ECCV 2008, Part I. LNCS, vol. 5302, pp. 497–508. Springer, Heidelberg (2008)
4. Rueda, A., Malpica, N., Romero, E.: Single-image super-resolution of brain MR images using overcomplete dictionaries. Med. Im. An. 17(1), 113–132 (2013)
5. Coupé, P., Manjón, J.V., Chamberland, M., Descoteaux, M., Hiba, B.: Collaborative patch-based super-resolution for diffusion-weighted images. NeuroImage 83, 245–261 (2013)
6. Ye, D.H., Zikic, D., Glocker, B., Criminisi, A., Konukoglu, E.: Modality propagation: coherent synthesis of subject-specific scans with data-driven regularisation. In: Mori, K., Sakuma, I., Sato, Y., Barillot, C., Navab, N. (eds.) MICCAI 2013, Part I. LNCS, vol. 8149, pp. 606–613. Springer, Heidelberg (2013)
7. Breiman, L.: Random forests. Machine Learning 45, 5–32 (2001)
8. Criminisi, A., Shotton, J.: Decision forests for computer vision and medical image analysis. Springer (2013)
9. Seunarine, K.K., Alexander, D.C.: Multiple fibers: beyond the diffusion tensor. In: Johansen-Berg, H., Behrens, T.E.J. (eds.) Diffusion MRI: from Quantitative Measurement to in Vivo Neuroanatomy, pp. 56–74. Academic Press (2009)