7T-Guided Learning Framework for Improving the Segmentation of 3T MR Images

Khosro Bahrami, Islem Rekik, Feng Shi, Yaozong Gao, and Dinggang Shen[✉]

Department of Radiology and BRIC,
University of North Carolina at Chapel Hill, Chapel Hill, NC, USA
dinggang_shen@med.unc.edu

Abstract. The emerging era of ultra-high-field MRI using 7T MRI scanners dramatically improved sensitivity, image resolution, and tissue contrast when compared to 3T MRI scanners in examining various anatomical structures. The advantages of these high-resolution MR images include higher segmentation accuracy of MRI brain tissues. However, currently, accessibility to 7T MRI scanners remains much more limited than 3T MRI scanners due to technological and economical constraints. Hence, we propose in this work the first learning-based model that improves the segmentation of an input 3T MR image with any conventional segmentation method, through the reconstruction of a higher-quality 7T-like MR image, without actually acquiring an ultra-high-field 7T MRI. Our proposed framework comprises two main steps. First, we estimate a non-linear mapping from 3T MRI to 7T MRI space, using random forest regression model with novel weighting and ensembling schemes, to reconstruct initial 7T-like MR images. Second, we use a group sparse representation with a new pre-selection approach to further refine the 7T-like MR image reconstruction. We evaluated our 7T MRI reconstruction results along with their segmentation results using 13 subjects acquired with both 3T and 7T MR images. For tissue segmentation, we applied two widely used segmentation methods (FAST and SPM) to perform the experiments. Our results showed (1) the improvement of WM, GM and CSF brain tissues segmentation results when guided by reconstructed 7T-like images compared to 3T MR images, and (2) the outperformance of the proposed 7T MRI reconstruction method when compared to other state-of-the-art methods.

1 Introduction

In the past years, a number of methods have been proposed to improve the segmentation of MR brain images into different tissues such as white matter (WM), gray matter (GM), and cerebrospinal fluid (CSF). However, besides the robustness and good accuracy of the used segmentation tools, the quality of the MR images highly affects image segmentation quality. Recently, numerous improvements have been proposed in strengthening the magnetic field of Magnetic Resonance Imaging (MRI) to obtain higher quality MR images. This has

led to the emergence of ultra-high-field (7T) MR images that yield better seg-
mentation results in comparison to the routine 3T MRI, as can be observed in
Fig. 1. This can be explained by the increase in the resolution of 7T MRI com-
pared to 3T MRI. Also, 7T MRI has higher sensitivity to tissue changes and
anatomical details, thereby generating higher tissue contrast and clearer WM,
GM, and CSF tissue boundaries [1, 2]. However, 7T MRI scanners are currently
more expensive and less available in hospitals and clinical centers, and thus the
majority of MR images are still acquired using routine 3T MRI scanners. In this
paper, we take on a new perspective for improving brain tissue segmentation by
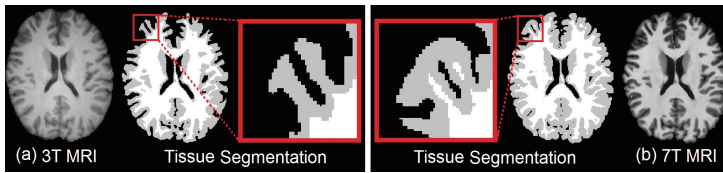reconstruction of 7T-like images from 3T MR images.

Fig. 1. (a) 3T MRI and (b) 7T MRI of the same subject, together with their segmen-
tation results. 7T shows better segmentation quality when compared with 3T.

The recent advances in the development of learning-based sparse representa-
tion methods result in the reconstruction of high-resolution (HR) images from
low-resolution (LR) images using paired LR and HR dictionaries, so that the
high-frequency details lost in the LR image can be predicted from the corre-
sponding HR dictionaries [3]. HR reconstruction from different image modalities
was also investigated in [4], where they introduced a MR image example-based
contrast synthesis (MIMECS) approach based on sparse representation. How-
ever, the learning-based sparse representation methods assume a high correla-
tion between LR and HR images, which may not be completely valid in practice.
To address this problem, Bahrami *et al.* [5] proposed multi-level canonical cor-
relation analysis (CCA) transform, called M-CCA, to increase the correlation
between LR and HR spaces. However, as a linear transformation, CCA may not
necessarily capture the non-linear nature of the mapping from LR to HR space.

In our paper, to overcome the problem of the large distribution gap between
LR (3T) and HR (7T) spaces, we propose a two-step framework. In the first
step, we generate a non-linear mapping from 3T to 7T MRI using random forest
(RF) regression to produce the initial 7T-like images with higher correlation and
similarity to the ground-truth 7T MRI. However, solely relying on RF non-linear
mapping may lead to fuzzy results due to the averaging operation across different
trees of RF. So, in the second step, we use a sparse representation between
the initial 7T-like images (i.e., outputted by the RF) and the ground-truth 7T
MRI to produce the final high-quality 7T-like image, which can be inputted to
any segmentation algorithm for tissue segmentation. Our key contributions are
as follows: (1) We proposed a framework, which capitalizes on a high-quality

reconstruction of 7T-like image, for improving 3T MR image segmentation. (2) We combined RF non-linear mapping with sparse representation to reconstruct the 7T-like images in two steps. Such framework addresses the flaws of both RF and SR. Particularly, the main flaw of RF's mapping is that it causes fuzzy results, due to the built-in averaging of the 7T patches. The key flaw of SR resides in the estimated sparse coefficients in the 3T space that are then directly applied to the 7T space; however, this may not be true due to large distribution gap between 3T and 7T image spaces. (3) For the RF, we introduced a weighting scheme by assigning higher weights to the more representative training samples and also a new ensembling strategy based on the distribution of the outputs of the training trees to remove outliers and avoid unreliable RF contributions. (4) For the sparse representation, we proposed a novel pre-selection scheme and also incorporated group-sparsity to share the sparsity among the outputs of RF regressions for a more reliable and accurate 7T-like image reconstruction.

2   Materials and Methods

2.1   MRI Dataset and Preprocessing

We used a dataset of 13 subjects acquired with both 3T and 7T MR images, with respective resolutions of $1 \times 1 \times 1$ mm^3 and $0.65 \times 0.65 \times 0.65$ mm^3. First, we corrected the bias field of each image followed by performing the skull stripping to remove non-brain tissues. Afterwards, all 7T images were linearly aligned to the MNI space guided by an individual template. Then, each 3T MRI was rigidly aligned to its corresponding 7T MRI.

2.2   Method

Figure 2 shows the conventional segmentation (top) *v.s.* our proposed method (bottom), which includes (1) the initial RF-based reconstruction, and (2) the sparse reconstruction-based refinement steps. In the following sections, these steps are explained in details.

2.2.1   RF-based Initial Reconstruction

In this step, the objective is to produce the initial 7T-like images with higher correlation and similarity to the ground-truth 7T MRI. Using the dataset of $N = 13$ pairs of 3T and 7T MR images, we use a leave-one-out strategy to use one pair for testing, and the remaining $N - 1 = 12$ pairs for the training. First, using the training set, we generate non-linear mappings between the training 3T MR and 7T MR images based on random forest (RF). To do so, we apply the leave-one-out strategy to the training dataset ($N - 1$ images) to learn a mapping for each training 3T image by using the remaining $N - 2$ training images. Each resulted mapping is used to generate the initial 7T-like training image from each 3T training image. Finally, we respectively apply the learned mappings to the testing 3T image to construct $N - 1$ initial 7T-like testing images.
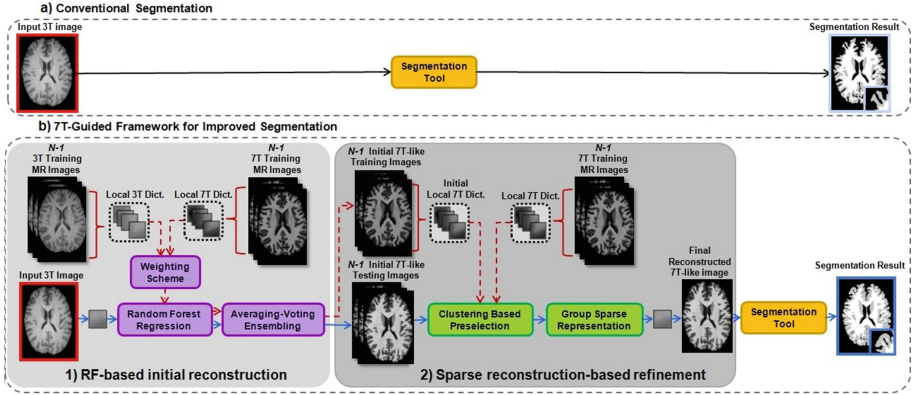
Fig. 2. (a) Conventional segmentation. (b) Our proposed 7T-Guided learning framework for the reconstruction of 7T-like images from 3T MRI to improve segmentation. In RF-based initial reconstruction, a leave-one-out cross-validation on the $N-1$ training images generates $N-1$ learned mappings and also $N-1$ initial 7T-like training images. The $N-1$ learned mappings are also applied to each input 3T image to generate $N-1$ initial 7T-like testing images. Red dashed and blue solid arrows indicate the testing and training steps, respectively.

RF has been widely used for classification and regression problems [6]. In this work we use RF for regression by mapping the input feature space onto the output target value space, where each input feature vector represents a vectorized 3T MRI patch and the output target value represents the voxel intensity in the corresponding 7T MRI patch. To map the 3T onto the 7T space at a given voxel location, we need to construct the 3T and 7T dictionaries. We define $\mathbf{x}$ to be a column vector representing a 3T patch of size $m \times m \times m$ extracted from the input 3T MR image $\mathbf{X}$. Then, for each 3T patch $\mathbf{x}$, we build both local 3T and 7T dictionaries ($\mathbf{D}_{3T}$ and $\mathbf{D}_{7T}$) by extracting the overlapped patches at the same location and also neighboring locations within a search window of size W from all $N-1$ pairs of aligned 3T and 7T training MR images.

Weighting Scheme. In the traditional RF regression, all the patches in the local dictionary are assumed to equally contribute to the generation of the mapping. However, in practice, the patches in the local dictionary may not have the same importance in the estimation of the mapping from 3T to 7T images. Here, we propose a *weighting scheme* to address this problem. Let $\mathbf{u} \in \mathbf{D}_{3T}$ be an input feature vector and $\mathbf{v} \in \mathbf{D}_{7T}$ its corresponding target values for regression. RF regression generates a mapping from the atoms of $\mathbf{D}_{3T}$ to the atoms of $\mathbf{D}_{7T}$. Hence, we define a weighting scheme for the patches in the local dictionary based on their similarity to the input 3T patch. To calculate the weights, we define an exponential function using L_2-norm distance between the input 3T patch $\mathbf{x}$ and each atom $\mathbf{u}$ in $\mathbf{D}_{3T}$, as $w = \exp(-\frac{||\mathbf{x}-\mathbf{u}||_2^2}{h})$, where h is used for adjusting the weight decay speed and is calculated based on the variance of the distribution

of the similarity between the input 3T patch $\mathbf{x}$ and each atom $\mathbf{u}$. To apply this weighting scheme, first, for each atom $\mathbf{u}$ in $\mathbf{D}_{3T}$, we normalize the weight w to be an integer in the range of [0, 5]. Then, proportionally to the weight, we replicate the training patches in the dictionary. In such case, the patches with higher similarity to the input patch will be largely replicated in the trees, thereby highly affecting the final RF result.

Averaging-Voting Ensembling. In the RF, the feature vector of each 3T intensity patch passes through each tree, and finally reaches one particular leaf node per decision tree. To derive the final prediction result, the traditional approach averages all the training samples in the arrived leaf nodes across all decision trees. However, this ensemble method may not be sufficiently robust and accurate, especially when there are some outliers or large variations in the outputs of different decision trees. To address this problem, we propose a new ensemble model, called *averaging-voting ensembling*, to reduce the effect of the outliers and unreliable results. Specifically, in our approach, we first split all T trees into G groups $T_1, ..., T_G$ using K-means clustering per patch. To do so, the feature vector of the 3T image patch passes through each tree, and reaches one particular leaf node per decision tree. We use the target value of that leaf node as a feature to categorize trees into groups. Then, for each group g $(g = 1, ..., G)$, we estimate its output initial 7T-like intensity value, as $y_g = \frac{1}{\sum_{t \in T_g} L_t} \sum_{t \in T_g} \sum_{l \in \{1, ..., L_t\}} v_l^t$, where v_l^t denotes the 7T MRI voxel value of the l-th training sample contained in the arrival leaf node of decision tree $t \in T_g$ of the g-th group, and L_t is the number of training samples contained in the arrival leaf node of the decision tree t. Next, we generate a histogram with B bins to estimate the distribution of the outputted values by all groups of decision trees. Finally, we output 7T-like intensity value by the bin with maximum count, which denotes the value that the majority of groups of trees produced.

2.2.2   Sparse Reconstruction-Based Refinement

Using the generated $N - 1$ initial 7T-like testing images by RF, the goal in this step is to reconstruct the final 7T-like image from all initial 7T-like testing images based on group sparsity. Hence, we build a local dictionary $\mathbf{D}'_{7T}$ for each patch by extracting the patches from all the initial 7T-like training images using a search window around the same location.

Clustering-based Pre-selection. In the local dictionary, all patches may not necessarily have the same importance in representing the input 3T patch. To select the best patches, we propose a method based on K-means clustering to divide the local dictionary $\mathbf{D}'_{7T}$ into P local sub-dictionaries $\mathbf{D}'_{7T,p}$ $(p = 1, ..., P)$ by minimizing the intra-cluster variance. In our method, we use both LR and HR patches as input features for clustering the local dictionary into local sub-dictionaries. In such case, we exploit the high-quality 7T patches to improve the clustering results. Correspondingly, we divide $\mathbf{D}_{7T}$ into P local sub-dictionaries $\mathbf{D}_{7T,p}$ $(p = 1, ..., P)$ using the same indices as $\mathbf{D}'_{7T,p}$. In this way, P paired local sub-dictionaries $\{\mathbf{D}'_{7T,p}, \mathbf{D}_{7T,p}\}_{p=1}^{P}$ are established. Then, the local

sub-dictionary whose elements have the minimum L_2-norm distance to the initial 7T-like patch is selected for the sparse representation of the initial 7T-like patch, denoted as $\mathbf{D}'_{7T,p_{min}}(p_{min} \in \{1,...,P\})$. Different from the previous pre-selection methods, which often use only the similarity of the training LR patches to the input LR patch, we benefit from the high-quality HR patches for better pre-selection by using both LR patches and HR patches as input features to cluster each local dictionary into local sub-dictionaries.

Group Sparse Representation. In the case of using sparse representation, each initial 7T-like image patch denoted as $\mathbf{y}$ can be sparsely represented using the local sub-dictionary $\mathbf{D}'_{7T,p_{min}}$ via $\hat{\boldsymbol{\alpha}} = \arg \min ||\mathbf{y} - \mathbf{D}'_{7T,p_{min}}\boldsymbol{\alpha}||_2^2 + \lambda||\boldsymbol{\alpha}||_1$, where $p_{min} \in \{1,...,P\}$ as mentioned above and $\hat{\boldsymbol{\alpha}}$ is the column vector of the sparse coefficients. Then, the estimated $\hat{\boldsymbol{\alpha}}$ can be utilized to reconstruct the final 7T-like patch, as $\mathbf{z} = \mathbf{D}_{7T,p_{min}}\hat{\boldsymbol{\alpha}}$. Different from the previous methods, here we incorporate $N-1$ initial 7T-like testing images, denoted as $\mathbf{Y}_i(i = 1,...,N-1)$, to reconstruct the final 7T-like image $\mathbf{Z}$. To do so, instead of sparsely representing every initial 7T-like patch independently, we incorporate the group sparsity using all initial 7T-like testing patches to allow them to share the same sparsity, thereby making the local structure consistent for the reconstructed patches. We reformulate the sparse representation to have group sparsity as $\hat{\mathbf{A}} = \arg \min \sum_{i=1}^{N-1} ||\mathbf{y}_i - \mathbf{D}'_{7T,p_{min},i}\boldsymbol{\alpha}_i||_2^2 + \lambda||\mathbf{A}||_{2,1}$, where the first term is a multi-task least square minimizer for $N-1$ patches from $N-1$ initial 7T-like testing images $\mathbf{Y}_i$, with $\mathbf{D}'_{7T,p_{min},i}$, $\mathbf{y}_i$, and $\boldsymbol{\alpha}_i$ denoting the initial local 7T sub-dictionary, a patch of the image $\mathbf{Y}_i$, and sparse coefficient vector of the i-th patch in the group, respectively. The second term is a regularizer, which is a combination of L_1 and L_2 norms on $\hat{\mathbf{A}} = [\boldsymbol{\alpha}_1,...,\boldsymbol{\alpha}_{N-1}]$, where each column of $\hat{\mathbf{A}}$ includes the sparse coefficients of a patch in the group. The L_2 norm is imposed on each row of $\hat{\mathbf{A}}$ for making the patches at the same location have similar sparsity, while imposing the L_1 norm on all rows of $\hat{\mathbf{A}}$ to make them sparse.

3   Experimental Results

We evaluated our method for the 7T-like image reconstruction and segmentation of the reconstructed 7T-like images into WM, GM, and CSF tissues. We compared the segmentation results based on 3T MRI and the reconstructed 7T-like images by different methods, including histogram-based, sparse representation [3], random forest [6], MIMECS [4] and M-CCA [5] and our method. We employed widely-used FAST in FSL package [7] and SPM [8] for tissue segmentation, where the segmentation result of 7T MR images is considered as ground-truth. For all the methods, we evaluated and reported the best results based on a grid search on the parameters. E.g., we used 100 trees in the random forest regression. For the averaging-voting ensembling, we used the K-means clustering with $G = 10$ clusters and the histogram with $B = 5$ bins. For sparse representation, we set $\lambda = 0.1$ and the K-means clustering with $P = 5$ clusters
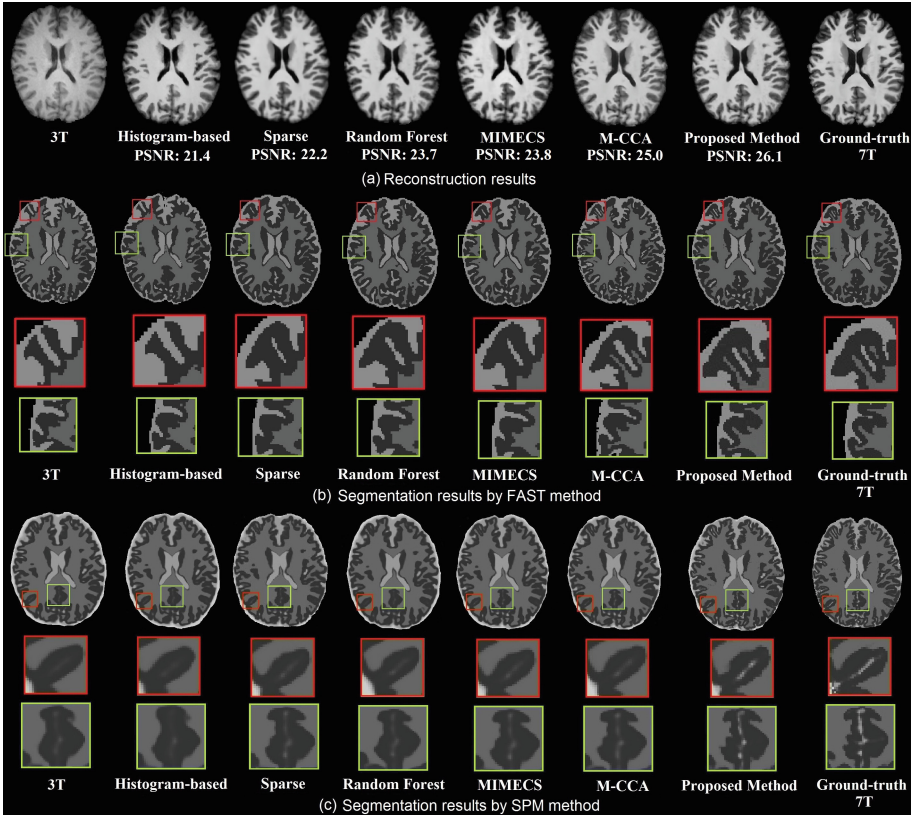
Fig. 3. (a) Visual and numerical (average PSNR across subjects) comparison of the reconstructed 7T-like MR images by different methods. (b) and (c) are segmentations of 3T, the reconstructed 7T-like MR images by different methods, and ground-truth 7T MRI by FAST and SPM methods, respectively.
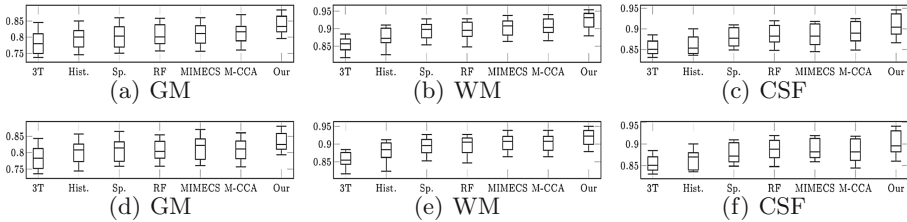

Fig. 4. Box plot of the average Dice ratio of (a)-(c) FAST and (d)-(f) SPM methods in segmentation of 3T MRI and also the reconstructed 7T-like MR images by histogram-based, sparse representation, Random Forest, MIMECS, M-CCA and our method.

for dividing the local dictionary into local sub-dictionary. We chose a patch size of $5 \times 5 \times 5$ and a search window size of $9 \times 9 \times 9$.

Figure 3 compares the reconstruction and segmentation results of our method and the previous methods. Compared to the previous methods, our method has better visual and numerical reconstruction results as reflected by the average PSNR value across subjects (Fig. 3 (a)). We evaluated the impact of RF (Random Forest), SR (Sparse Representation), WI (Weighted Input), AV (Averaging-Voting), CP (Clustered-based Pre-selection), GS (Group Sparsity) on performance of our method. The performance improvement in terms of average PSNR includes, RF+SR: 24.9; RF+SR+WI: 25.2; RF+SR+WI+AV: 25.6; RF+SR+WI+AV+CP: 25.8; RF+SR+WI+AV+CP+GS: 26.1, which reveals the importance of the proposed contributions.

Figure 3 (b) and (c) display the segmentation results using FAST and SPM, respectively, together with the close-up views of selected regions, for 3T MRI, the reconstructed 7T-like images by different methods, and the ground-truth 7T MRI. Compared to 3T MRI, the segmentation result by our method is much closer to the segmentation of the ground-truth 7T MRI. Also, our method has better WM, GM, and CSF brain tissue segmentation results than those using other 7T MRI reconstruction methods. For a quantitative comparison, the average Dice ratio of segmentation maps for WM, GM, and CSF for all 13 subjects based on FAST and SPM methods are displayed in Fig. 4 (a)-(c) and (d)-(f), respectively. Generally, our method has superior results compared to the previous reconstruction methods ($p < 0.01$ by two-sample test), and our segmentation results outperformed the direct segmentation results from 3T MRI.

4   Conclusion

We proposed a learning framework for reconstructing 7T-like MR image from 3T MRI to improve the segmentation accuracy of brain tissues using any conventional segmentation method. The experimental results showed that our proposed method outperformed the previous 7T-like MRI reconstruction methods both visually and numerically. Furthermore, our reconstructed 7T-like MR images led to significantly higher accuracy of WM, GM, and CSF brain tissue segmentations compared to directly using 3T MRI. Although we used FAST and SPM for segmentation, one could combine our framework with any segmentation method for better segmenting tissues from 3T MRI.

References

1. Kolka, A.G., Hendriksea, J., Zwanenburg, J.J.M., Vissera, F., Luijtena, P.R.: Clinical applications of 7T MRI in the brain. European Journal of Radiology **82**, 708–718 (2013)
2. Beisteiner, R., Robinson, S., Wurnig, M., Hilbert, M., Merksa, K., Rath, J., Hllinger, I., Klinger, N., Marosi, C., Trattnig, S., Geiler, A.: Clinical fMRI: Evidence for a 7T benefit over 3T. NeuroImage **57**, 1015–1021 (2011)
3. Yang, J., Wright, J., Huang, T., Ma, Y.: Image super-resolution via sparse representation. IEEE. Trans. Image Process. **19**(11), 2861–2873 (2010)

4. Roy, S., Carass, A., Prince, J.L.: Magnetic Resonance Image Example Based Contrast Synthesis. IEEE Trans. Med. Imaging **32**, 2348–2363 (2013)
5. Bahrami, K., Shi, F., Zong, X., Shin, H.W., An, H., Shen, D.: Hierarchical reconstruction of 7T-like images from 3T MRI using multi-level CCA and group sparsity. In: Navab, N., Hornegger, J., Wells, W.M., Frangi, A.F. (eds.) MICCAI 2015. LNCS, vol. 9350, pp. 659–666. Springer, Heidelberg (2015). doi:10.1007/978-3-319-24571-3_79
6. Alexander, D.C., Zikic, D., Zhang, J., Zhang, H., Criminisi, A.: Image quality transfer via random forest regression: applications in diffusion MRI. In: Golland, P., Hata, N., Barillot, C., Hornegger, J., Howe, R. (eds.) MICCAI 2014. LNCS, vol. 8675, pp. 225–232. Springer, Heidelberg (2014). doi:10.1007/978-3-319-10443-0_29
7. Zhang, Y., Brady, M., Smith, S.: Segmentation of brain MR images through Markov random field model and the expectation-maximization algorithm. IEEE Trans. Med. Imag. **20**(1), 45–57 (2001)
8. Ashburner, J., Friston, K.J.: Unified segmentation. NeuroImage **29**, 839–851 (2005)