Tree-Encoded Conditional Random Fields for Image Synthesis

Amod Jog$^{1(\boxtimes)}$, Aaron Carass^1, Dzung L. Pham^2, and Jerry L. Prince^1

^{1} Image Analysis and Communications Laboratory,
The Johns Hopkins University, Baltimore, USA
amodjog@jhu.edu
^{2} Henry M. Jackson Foundation for the Advancement of Military Medicine,
Bethesda, USA

Abstract. Magnetic resonance imaging (MRI) is the dominant modality for neuroimaging in clinical and research domains. The tremendous versatility of MRI as a modality can lead to large variability in terms of image contrast, resolution, noise, and artifacts. Variability can also manifest itself as missing or corrupt imaging data. Image synthesis has been recently proposed to homogenize and/or enhance the quality of existing imaging data in order to make them more suitable as consistent inputs for processing. We frame the image synthesis problem as an inference problem on a 3-D continuous-valued conditional random field (CRF). We model the conditional distribution as a Gaussian by defining quadratic association and interaction potentials encoded in leaves of a regression tree. The parameters of these quadratic potentials are learned by maximizing the pseudo-likelihood of the training data. Final synthesis is done by inference on this model. We applied this method to synthesize T_2-weighted images from T_1-weighted images, showing improved synthesis quality as compared to current image synthesis approaches. We also synthesized Fluid Attenuated Inversion Recovery (FLAIR) images, showing similar segmentations to those obtained from real FLAIRs. Additionally, we generated super-resolution FLAIRs showing improved segmentation.

Keywords: Magnetic resonance · Image synthesis · Conditional random field

1 Introduction

Image synthesis in MRI is a process in which the intensities of acquired MRI data are transformed in order to enhance the data quality or render them more suitable as input for further image processing. Image synthesis has been gaining traction in the medical image processing community in recent years [6,15], as a useful pre-processing tool for segmentation and registration. It is especially useful in MR brain imaging, where a staggering variety of pulse sequences like Magnetization Prepared Gradient Echo (MPRAGE), Dual Spin Echo (DSE),

FLAIR etc. are used to interrogate the various aspects of neuroanatomy. Versatility of MRI is a boon for diagnosticians but can sometimes prove to be a handicap when performing analysis using image processing. Automated image processing algorithms are not always robust to variations in their input [13]. In large datasets, sometimes images are missing or corrupted during acquisition and cannot be used for further processing. Image synthesis can be used as a tool to supplement these datasets by creating artificial facsimiles of the missing images by using the available ones. An additional source of variability is the differing image quality between different pulse sequences for the same subject. An MPRAGE sequence can be quickly acquired at a resolution higher than $1\,\mathrm{mm}^3$, which is not possible for FLAIR. Image synthesis can be used to enhance the resolution of existing low resolution FLAIRs using the corresponding high resolution MPRAGE images, thus leading to improved tissue segmentation.

Previous work on image synthesis has proceeded along two lines, (1) registration-based, and (2) example-based. Registration-based [3,12] approaches register the training/atlas images to the given subject image and perform intensity fusion (in the case of multiple training/atlas pairs) to produce the final synthesis. These approaches are heavily dependent on the quality of registration, which is generally not accurate enough in the cortex and abnormal tissue regions. Example-based approaches involve learning an intensity transformation from known training data pairs/atlas images. A variety of example-based approaches [5,8,15] have been proposed. These methods treat the problem as a regression problem and estimate the synthetic image voxel-by-voxel from the given, available images. The voxel intensities in the synthetic image are assumed to be independent of each other, which is not entirely valid as intensities in a typical MR image are spatially correlated and vary smoothly from voxel-to-voxel.

In this work, we frame image synthesis as an inference problem in a probabilistic discriminative framework. Specifically, we model the posterior distribution $\mathrm{p}(\mathbf{y}|\mathbf{x})$, where $\mathbf{x}$ is the collection of known images and $\mathbf{y}$ is the synthetic image we want to estimate, as a Gaussian CRF [10]. Markov random field (MRF) approaches lend themselves as a robust, popular way to model images. However in a typical MRF, the observed data $\mathbf{x}$ are assumed to be independent given the underlying latent variable $\mathbf{y}$, which is a limiting assumption for typical images. A CRF, by directly modeling the posterior distribution allows us to side-step this problem. CRFs have been used in discrete labeling and segmentation problems [9]. A continuous-valued CRF, modeled as a Gaussian CRF, was first described in [17]. Efficient parameter learning and inference procedures of a Gaussian CRF were explored in the Regression Tree Fields concept in [7]. We also model the posterior distribution as a Gaussian CRF, the parameters of which are stored in the leaves of a single regression tree. We learn these parameters by maximizing a pseudo-likelihood objective function given training data. Given a subject image, we build the Gaussian distribution parameters from the learned tree and parameters. The prediction of the synthetic subject image is a maximum a posteriori (MAP) estimate of this distribution and is estimated efficiently using conjugate gradient descent.

We refer to our method as Synthesis with <u>C</u>onditional <u>R</u>andom <u>F</u>ield <u>T</u>ree or SyCRAFT. We applied SyCRAFT to synthesize T_2-weighted (T_2-w) images from T_1-w (T_1-w) images and showed a superior quality of synthesis compared to state-of-the-art methods. We also applied our method to synthesize FLAIRs from corresponding T_1-w, T_2-w, and P_D-weighted (P_D-w) images and showed that tissue segmentation on synthetic images is comparable to that achieved using real images. Finally, we used our method in an example-based super-resolution framework to estimate a super-resolution FLAIR image and showed improved tissue segmentation. In Sect. 2, we describe our method in detail, followed by experiments and results in Sect. 3 and discussion in Sect. 4.

2   Method

2.1   Model

We start with the definition of a CRF, initially proposed in [10]. A CRF is defined over a graph $G = (V, E)$, V and E are the sets of vertices and edges respectively, of G. In an image synthesis context, the set of all voxels i in the image domain form the vertices of V. A pair of voxels $(i, j), i, j \in V$, that are neighbors according to a predefined neighborhood, form an edge in E. Let $\mathbf{x} = \{\mathbf{x}_1, \ldots, \mathbf{x}_m\}$ be the observed data. Specifically, $\mathbf{x}$ represents the collection of available images from m pulse sequences from which we want to synthesize a new image. Let $\mathbf{y}$ be the continuous-valued random variable over V, representing the synthetic image we want to predict. In a CRF framework, $\mathrm{p}(\mathbf{y}|\mathbf{x})$ is modeled and learned from training data of known pairs of $(\mathbf{x}, \mathbf{y})$. Let $\mathbf{y} = \{y_i, i \in V\}$. Then $(\mathbf{y}, \mathbf{x})$ is a CRF if, conditioned on $\mathbf{x}$, y_i exhibit the Markov property, i.e. $\mathrm{p}(y_i|\mathbf{x}, \mathbf{y}_{V \setminus i}) = \mathrm{p}(y_i|\mathbf{x}, \mathbf{y}_{\mathcal{N}_i})$, where $\mathcal{N}_i = \{j \mid (i, j) \in E\}$, is the neighborhood of i.

Assuming $\mathrm{p}(\mathbf{y}|\mathbf{x}) > 0, \forall \mathbf{y}$, from the Hammersley-Clifford theorem, we can express the conditional probability as a Gibbs distribution. The factorization of $\mathrm{p}(\mathbf{y}|\mathbf{x})$ in terms of association potentials and interaction potentials is given as,

$$\mathrm{p}(\mathbf{y}|\mathbf{x}) = \frac{1}{Z} \exp[-\{\sum_{i \in V} E_{\mathcal{A}}(y_i, \mathbf{x}; \theta) + \lambda \sum_{i \in V} \sum_{j \in \mathcal{N}_i} E_{\mathcal{I}}(y_i, y_j, \mathbf{x}; \theta)\}]. \qquad (1)$$

$E_{\mathcal{A}}(y_i, \mathbf{x}; \theta)$ is called an association potential, defined using the parameter set θ, $E_{\mathcal{I}}(y_i, y_j, \mathbf{x}; \theta)$ is called an interaction potential, λ is a weighting factor, and Z is the partition function. If E_A and E_I are defined as quadratic functions of $\mathbf{y}$, we can express this distribution as a multivariate Gaussian as below,

$$\mathrm{p}(\mathbf{y}|\mathbf{x}) = \frac{1}{(2\pi)^{\frac{|V|}{2}} |\boldsymbol{\Sigma}|^{\frac{1}{2}}} \exp(-\frac{1}{2}(\mathbf{y} - \boldsymbol{\mu}(\mathbf{x}))^T \boldsymbol{\Sigma}(\mathbf{x})^{-1}(\mathbf{y} - \boldsymbol{\mu}(\mathbf{x})))$$
$$= \frac{1}{Z} \exp(-\frac{1}{2}(\mathbf{y}^T \mathbf{A}(\mathbf{x})\mathbf{y}) - \mathbf{b}(\mathbf{x})^T \mathbf{y}). \qquad (2)$$

The parameters $\mathbf{A}(\mathbf{x})$ and $\mathbf{b}(\mathbf{x})$ are dependent on the association and interaction potential definitions. In most classification tasks involving CRF's the association

potential is defined as the local class probability as provided by a generic classifier or a regressor [9]. Image synthesis being a regression task, we chose to model and extract both association and interaction potentials from a single regressor, in our case a regression tree. We define a quadratic association potential as

$$E_{\mathcal{A}}(y_i, \mathbf{x}; \theta) = \frac{1}{2}(a_{L(i)}y_i^2) - b_{L(i)}y_i, \tag{3}$$

where $\{a_{L(i)}, b_{L(i)}\} \in \theta$ are the parameters defined at the leaf $L(i)$. $L(i)$ is the leaf where the feature vector $\mathbf{f}_i(\mathbf{x})$ extracted for voxel i from the observed data $\mathbf{x}$, lands after having been passed through successive nodes of a learned regression tree, Ψ. The features and regression tree construction are described in Sect. 2.2.

The interaction potential usually acts as a smoothing term, but can also be designed in a more general manner. We define interaction potentials for each type of neighbor. A 'neighbor type' $r \in \{1, \ldots, |\mathcal{N}_i|\}$ is given by a voxel i and one of its $n\ (= |\mathcal{N}_i|)$ neighbors. For example, a neighborhood system with four neighbors (up, down, left, right) has four types of neighbors, and hence four types of edges. The complete set of edges E can be divided into non-intersecting subsets $\{E_1, \ldots, E_r, \ldots, E_n\}$ of edges of different types. Let the voxel j be such that $(i, j) \in E$ is a neighbor of type r, that is $(i, j) \in E_r$. Let the corresponding feature vectors $\mathbf{f}_i(\mathbf{x})$ and $\mathbf{f}_j(\mathbf{x})$ land in leaves $L(i)$ and $L(j)$ of the trained tree Ψ, respectively. The interaction potential is modeled as

$$E_{\mathcal{I}}(y_i, y_j, \mathbf{x}; \theta) = \frac{1}{2}(\alpha_{L(i)_r}y_i^2 + \beta_{L(i)_r}y_iy_j + \gamma_{L(i)_r}y_j^2) - \omega_{L(i)_{1r}}y_i - \omega_{L(i)_{2r}}y_j. \tag{4}$$

Let the set of leaves of the regression tree Ψ be $\mathcal{L}_\Psi$. Each leaf $l \in \mathcal{L}_\Psi$ stores the set of parameters $\theta_l = \{a_l, b_l, \alpha_{l1}, \beta_{l1}, \gamma_{l1}, \omega_{l11}, \omega_{l21}, \ldots, \alpha_{ln}, \beta_{ln}, \gamma_{ln}, \omega_{l1n}, \omega_{l2n}\}$. The complete set of parameters is thus, $\theta = \{\theta_l | l \in \mathcal{L}_\Psi\}$. Our approach bears similarity to the regression tree fields concept introduced in [7], where the authors create a separate regression tree for each neighbor type. Thus with a single association potential and a typical 3D neighborhood of 26 neighbors, they would need 27 separate trees to learn the model parameters. Training a large number of trees with large training sets makes the regression tree fields approach computationally expensive. It was especially not feasible in our application with large 3D images, more neighbors, and high dimensional feature vectors. We can however train multiple trees using bagging to create an ensemble of models to create an average, improved prediction. The training of a single regression tree is described in the next section.

2.2 Learning a Regression Tree

As mentioned before, let $\mathbf{x} = \{\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_m\}$ be a collection of co-registered images, generated by modalities $\Phi_1, \ldots, \Phi_m$, respectively. The image synthesis task entails predicting the image $\mathbf{y}$ of a target modality Φ_t. The training data thus consists of known co-registered pair of $\{\mathbf{x}, \mathbf{y}\}$. At each voxel location i, we extract features $\mathbf{f}_i(\mathbf{x})$, derived from $\mathbf{x}$. For our experiments we use two types

of features, (1) small, local patches, (2) context descriptors. A small 3D patch, denoted by $\mathbf{p}_i(\mathbf{x}) = [\mathbf{p}_i(\mathbf{x}_1), \ldots, \mathbf{p}_i(\mathbf{x}_m)]$, where the size of the patch is typically $3 \times 3 \times 3$ and provides us with local intensity information.

We construct the context descriptors as follows. The brain images are rigidly aligned to the MNI coordinate system [4] with the center of the brain approximately at the center of the image. Thus for each voxel we can find out the unit vector $\mathbf{u}$ from the voxel i to the origin. We can define 8 directions by rotating the component of $\mathbf{u}$ in the axial plane by angles $\{0, \frac{\pi}{4}, \ldots, \frac{7\pi}{4}\}$. In each of these directions, we select average intensities of cubic regions of cube-widths $\{w_1, w_2, w_3, w_4\}$ at four different radii $\{r_1, r_2, r_3, r_4\}$ respectively. This becomes a 32-dimensional descriptor of the spatial context surrounding voxel i. In our experiments we used $w_1 = 3, w_2 = 5, w_3 = 7, w_4 = 9$ and $r_1 = 4, r_2 = 8, r_3 = 16, r_4 = 32$. These values were chosen empirically. We denote this context descriptor by $\mathbf{c}_i(\mathbf{x})$. The final feature vector is thus $\mathbf{f}_i(\mathbf{x}) = [\mathbf{p}_i(\mathbf{x}), \mathbf{c}_i(\mathbf{x})]$. $\mathbf{f}_i(\mathbf{x})$ is paired with the voxel intensity y_i at i in the target modality image $\mathbf{y}$ to create training data pairs $(\mathbf{f}_i(\mathbf{x}), y_i)$. We train the regression tree Ψ on this training data using the algorithm described in [2]. Once the tree is constructed, we initialize θ_l at each of the leaves $l \in \mathcal{L}_\Psi$. θ_l is estimated by a pseudo-likelihood maximization approach.

2.3   Parameter Learning

An ideal approach to learn parameters would be to perform maximum likelihood using the distribution in Eq. 2. However as mentioned in [7], estimation of the mean parameters $\boldsymbol{\Sigma}$ and $\boldsymbol{\mu}$, requires calculation of $\mathbf{A}^{-1}$ (see Eq. 2). The size of $\mathbf{A}$ is $|V \times V|$ where $|V|$ is the number of voxels in $\mathbf{y}$ and for large 3D images, $|V|$ is of the order of $\sim 10^6$, which makes the computation practically infeasible. We follow [7] and implement a pseudo-likelihood maximization-based parameter learning.

Pseudo-likelihood is defined as the product of local conditional likelihoods,

$$\hat{\theta}_{\mathrm{MPLE}} = \mathrm{argmax}_\theta \prod_{i \in V} \mathrm{p}(y_i \mid \mathbf{y}_{\mathcal{N}_i}, \mathbf{x}; \theta). \tag{5}$$

The local conditional likelihood can be expanded as

$$\mathrm{p}(y_i \mid \mathbf{y}_{\mathcal{N}_i}, \mathbf{x}; \theta) = \frac{\mathrm{p}(y_i, \mathbf{y}_{\mathcal{N}_i}, \mathbf{x}; \theta)}{\int_{\mathbb{R}} \mathrm{p}(y_i, \mathbf{y}_{\mathcal{N}_i}, \mathbf{x}; \theta) dy_i},$$

$$- \log \mathrm{p}(y_i \mid \mathbf{y}_{\mathcal{N}_i}, \mathbf{x}; \theta) = - \log \mathrm{p}(y_i, \mathbf{y}_{\mathcal{N}_i}, \mathbf{x}; \theta) + \log Z_i, \tag{6}$$

where $Z_i = \int_{\mathbb{R}} \mathrm{p}(y_i, \mathbf{y}_{\mathcal{N}_i}, \mathbf{x}; \theta) dy_i$. Using the CRF definition in Eq. 1, we can write $- \log \mathrm{p}(y_i, \mathbf{y}_{\mathcal{N}_i}, \mathbf{x}; \theta)$ as

$$- \log \mathrm{p}(y_i, \mathbf{y}_{\mathcal{N}_i}, \mathbf{x}; \theta) = E_{\mathcal{A}}(y_i, \mathbf{x}; \theta) + \lambda \sum_{j \in \mathcal{N}_i} E_{\mathcal{I}}(y_i, y_j, \mathbf{x}; \theta)$$

$$= \frac{1}{2} a_{Ci} y_i^2 - b_{Ci} y_i, \tag{7}$$

where we can find a_{Ci}(Eq. 8) and b_{Ci},(Eq. 9) by matching quadratic and linear terms. Equations 8 and 9 show the contribution of interaction potentials induced by the neighbors of voxel i. The $\widetilde{r}$ denotes the type of edge which is symmetric to type r. For example, if edges of type r are between voxel i and its right neighbor, then $\widetilde{r}$ denotes the type that is between a voxel and its left neighbor.

$$a_{Ci} = a_{L(i)} + \lambda \Big(\sum_{j|(i,j)\in E_r} \alpha_{L(i)_r} + \sum_{h|(h,i)\in E_{\widetilde{r}}} \gamma_{L(h)_{\widetilde{r}}} \Big) \tag{8}$$

$$b_{Ci} = b_{L(i)} + \lambda \Big(\sum_{j|(i,j)\in E_r} \omega_{L(i)_{1r}} + \sum_{h|(h,i)\in E_{\widetilde{r}}} \omega_{L(h)_{2\widetilde{r}}}$$
$$- \frac{1}{2} \sum_{j|(i,j)\in E_r} \beta_{L(i)_r} y_j - \frac{1}{2} \sum_{h|(h,i)\in E_{\widetilde{r}}} \beta_{L(h)_{\widetilde{r}}} y_h \Big). \tag{9}$$

The integral of exponential terms Z_i in Eq. 6, is also known as the log partition term. To optimize objective functions with log partition terms, we express Z_i in its variational representation using the mean parameters $\boldsymbol{\mu}_i = [\mu_i, \sigma_i]$ [18]. The parameter set $\boldsymbol{\theta}_{Ci} = \{b_{Ci}, a_{Ci}\}$ that defines the exponential distribution is known as the canonical parameter set. The conjugate dual function of Z_i is defined as follows,

$$Z_i^*(\mu_i, \sigma_i) = \sup_{\theta_{Ci}} \langle \boldsymbol{\theta}_{Ci}, \boldsymbol{\mu}_i \rangle - Z_i(\boldsymbol{\theta}_{Ci}), \tag{10}$$

where $\langle \rangle$ denotes inner product. Substituting $\boldsymbol{\theta}_{Ci}$ and the expression for $-\log \mathrm{p}(y_i, \mathbf{y}_{\mathcal{N}_i}, \mathbf{x}; \theta)$ from Eq. 7, we get the negative pseudo-likelihood contributed by voxel i to be,

$$\mathrm{NPL}_i(\theta) = b_{Ci}(\mu_i - y_i) + \frac{1}{2} a_{Ci}(y_i^2 - \sigma_i) + \log(\sigma_i - \mu_i^2) + \log(2\pi e), \tag{11}$$

where the mean parameters are given by $\mu_i = \frac{b_{Ci}}{a_{Ci}}$ and $\sigma_i = \frac{1}{a_{Ci}} + \mu_i^2$.

Equation 11 is similar to the one in [7], as the overall model is a Gaussian CRF. The objective function is linearly related to θ and is thus convex [7,18]. We minimize $\sum_{i\in V} \mathrm{NPL}_i(\theta)$ using gradient descent. The weighting factor $\lambda = 0.1$, was chosen empirically in our experiments. The regression tree fields concept performed a constrained, projected gradient descent on the parameters to ensure positive definiteness of the final precision matrix ($\mathbf{A}(\mathbf{x})$ in Eq. 1) [7]. We observed that unconstrained optimization in our model and applications generated a positive definite $\mathbf{A}(\mathbf{x})$. Training in our experiments takes about 20–30 min with $\sim 10^6$ samples of dimensionality of the order of $\sim 10^2$ and neighborhood size of 26, on a 12 core 3.42 GHz machine.

2.4   Inference

Given a test image set $\hat{\mathbf{x}} = \{\hat{\mathbf{x}}_1, \dots, \hat{\mathbf{x}}_m\}$, which are co-registered, we first extract features $\mathbf{f}_i(\hat{\mathbf{x}})$ from all voxel locations i. Next, we apply the learned regression
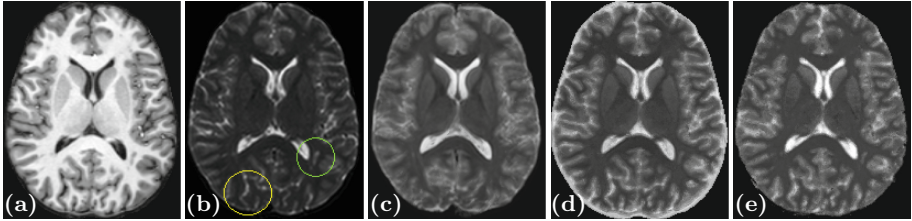
Fig. 1. Shown are **(a)** the input MPRAGE image, **(b)** the true T_2-w image, and the synthesis results from the MPRAGE for each of **(c)** FUSION, **(d)** MIMECS, and **(e)** SyCRAFT (our method). The lesion (green circle) and the cortex (yellow circle) in the true image are synthesized by MIMECS and SyCRAFT, but not by FUSION (Colour figure online).

tree Ψ to each of $\mathbf{f}_i(\hat{\mathbf{x}})$ to determine the leaf node $L(i)$ in Ψ. Using the learned parameters at these leaves, we construct the matrix $\mathbf{A}(\hat{\mathbf{x}})$ and the vector $\mathbf{b}(\hat{\mathbf{x}})$, (see Eq. 2). The diagonal and off-diagonal elements of $\mathbf{A}(\hat{\mathbf{x}})$ are populated by matching the linear and quadratic terms from Eq. 2. The MAP estimate for $p(\mathbf{y}|\hat{\mathbf{x}})$ as well as the conditional expectation $E[\mathbf{y}|\hat{\mathbf{x}}]$ is the mean of the multivariate Gaussian described in Eq. 2. The expression for the mean and hence the estimate $\hat{\mathbf{y}}$ is given by,

$$\hat{\mathbf{y}} = \mathbf{A}(\hat{\mathbf{x}})^{-1}\mathbf{b}(\hat{\mathbf{x}}). \tag{12}$$

$\mathbf{A}(\hat{\mathbf{x}})$ is a large ($\sim 10^6 \times 10^6$), sparse ($\sim 27 \times 10^6$ non-zero entries), symmetric positive definite matrix. Thus, we use an iterative preconditioned conjugate gradient descent method to solve the linear system in Eq. 12. The estimate $\hat{\mathbf{y}}$ is our synthetic image. Estimates from multiple (5 in our experiments) trained models using bagging can also be averaged to produce a final result.

3 Results

### 3.1 Synthesis of T_2-w Images from T_1-w Images

In this experiment, we used MPRAGE images from the publicly available multimodal reproducibility (MMRR) data [11] and synthesized the T_2-w images of the DSE sequence. The multimodal reproducibility data consists of 21 subjects, each with two imaging sessions, acquired within an hour of each other. Thus there are 42 MPRAGE images. We excluded data of five subjects (ten images), which were used for training and synthesized the remaining 32. We compared SyCRAFT to MIMECS [15] and multi-atlas registration and intensity fusion (FUSION) [3]. We used five subjects as the atlases for FUSION with the parameters $\beta = 0.5$ and $\kappa = 4$ (fuse the four best patch matches).

We used PSNR (peak signal to noise ratio), universal quality index (UQI) [19], and structural similarity (SSIM) [20] as metrics. UQI and SSIM take into account image degradation as observed by a human visual system. Both have values that lie between 0 and 1, with 1 implying that the images

Table 1. Mean and standard deviation (Std. Dev.) of the PSNR, UQI, and SSIM values for synthesis of T_2-w images from 32 MPRAGE scans.

	PSNR Mean (Std)	UQI Mean (Std.)	SSIM Mean (Std)
FUSION	52.73 (2.78)[a]	0.78 (0.02)	0.82 (0.02)
MIMECS	36.13 (2.23)	0.78 (0.02)	0.77 (0.02)
SyCRAFT	49.73 (1.99)	0.86 (0.01)[a]	0.84 (0.01)[a]

[a] Statistically significantly better than either of the other two methods (α level of 0.01) using a right-tailed test.

Subject T_1-w	T_2-w	P_D-w	Synth. FLAIR	True FLAIR
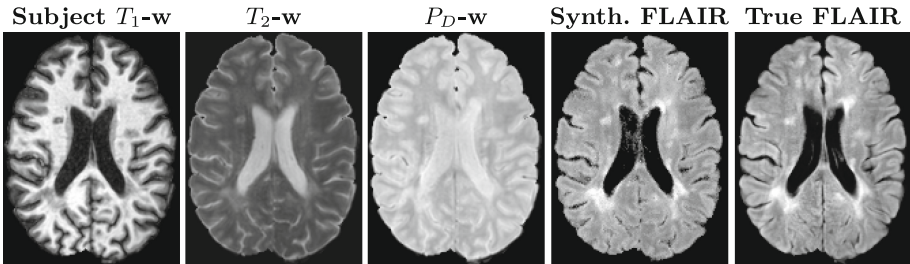
Fig. 2. Subject input images along with the SyCRAFT FLAIR and true FLAIR images.

are equal to each other. SyCRAFT performs significantly better than both the methods for all metrics except PSNR. Figure 1 shows the results for all three methods along with the true T_2-w image. FUSION results (Fig. 1(b)) have the highest PSNR, but produce anatomically incorrect images, especially in the presence of abnormal tissue anatomy (lesions for example) and the cortex. Overall, SyCRAFT produces an image that is visually closest to the true T_2-w image (Table 1).

3.2 Synthesis for FLAIR Images

In this experiment, given atlas P_D-w, T_2-w, T_1-w, and FLAIR images, we trained SyCRAFT and applied it to subject P_D-w , T_2-w, and T_1-w images, to predict the subject synthetic FLAIR image. We used our in-house multiple sclerosis (MS) patient image dataset with 49 subject images, with four training subjects and testing on the remaining 45. We computed average PSNR (20.81, std = 1.19), UQI (0.81, std = 0.03) and SSIM (0.78, std = 0.03), over these 45 subjects. These values indicate that the synthetic FLAIRs are structurally and visually similar to their corresponding real FLAIRs. Figure 2 shows the input images and the synthetic FLAIR image along side the real FLAIR image.

Next, we investigated the segmentations acquired from these synthetic FLAIRs. We would like the segmentation algorithm, LesionTOADS [16], to behave identically for real and synthetic images. LesionTOADS uses a T_1-w and
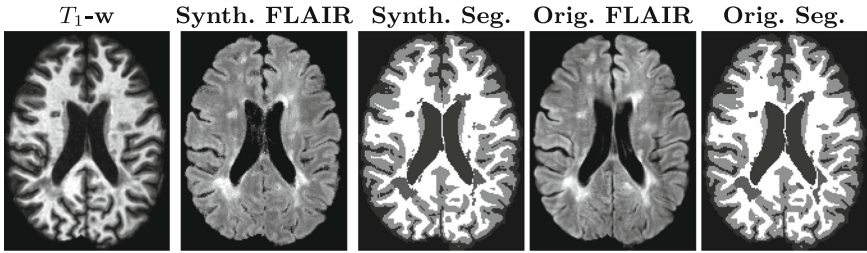
Fig. 3. LesionTOADS segmentations for real and synthetic FLAIRs.
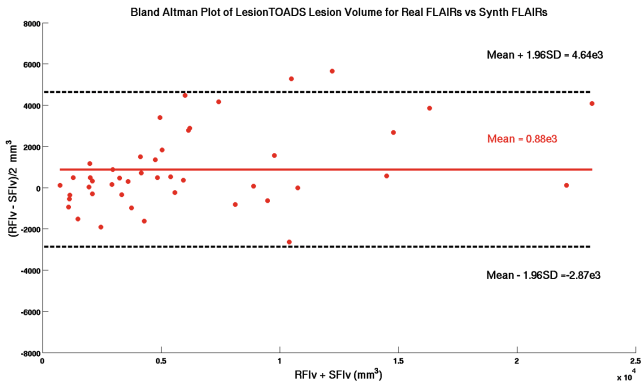

Fig. 4. A Bland-Altman plot of lesion volumes for synthetic FLAIRs vs real FLAIRs.

a corresponding FLAIR to generate a multi-class, topologically correct segmentation in the presence of lesions. We compared the overlap of segmentations obtained using synthetic FLAIRs to those obtained using real FLAIRs in terms of Dice coefficients (averaged over 45 subjects) for white matter (0.97, std = 0.01) (WM), gray matter (0.99, std = 0.01) (GM), cerebrospinal fluid (0.97, std = 0.01) (CSF) and white matter lesions (0.52, std = 0.17) (WML). Figure 3 shows the segmentations by LesionTOADS on real and synthetic FLAIRs. The overlap is very good for WM, GM, and CSF, however it is low for the WML class. The lesions being small and diffuse, even a small difference in the overlap can cause a low value for the Dice coefficient. So, we looked at the overall lesion volumes as provided by the algorithm for real and synthetic FLAIRs. To understand how different the lesion volumes are for the synthetic images as compared to the real images, we created a Bland-Altman [1] plot shown in Fig. 4. Let RFlv be the lesion volumes given by LesionTOADS using real FLAIRs as input. Let SFlv be the lesion volumes using synthetic FLAIRs as input. Bland-Altman plot is a scatter plot of $(RFlv - SFlv)/2$ (y axis) vs. $RFlv + SFlv$ (x axis). The measurements are considered to be interchangeable if 0 lies within $\pm 1.96\sigma$ where σ is the standard deviation of $(RFlv - SFlv)/2$. There is a small bias between RFlv and SFlv (mean =

0.88×10^3) however, 0 does lie between the prescribed limits and hence based on this plot we can say that these two measurements are interchangeable.

3.3   Super-Resolution of FLAIR

Next, we applied SyCRAFT to synthesize super-resolution (SR) FLAIRs using corresponding high resolution (HR) MPRAGE and low resolution (LR) FLAIRs. During a clinical or a research scan, not all the pulse sequences acquired are acquired at the same fixed resolution. Sequences like the T_1-w MPRAGE can be acquired very fast and hence are easy to image at a high resolution–usually higher than $1\,mm^3$ isotropic. However sequence like DSE and FLAIR have long repetition times (TR) and inversion times (TI), which limits the amount of scan time, and therefore, are acquired at a low (2–5 mm) through plane resolution.
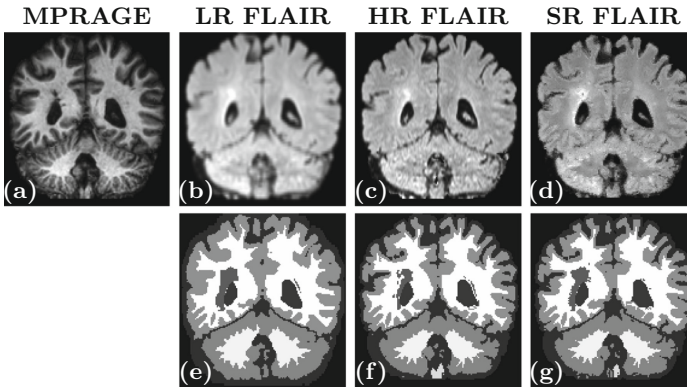

Fig. 5. Coronal slices of LR, HR, and SR FLAIRs along with their corresponding LesionTOADS segmentation are shown. It is evident that using a LR FLAIR affects the segmentation of the lesions and even the cortex.

Our approach can be described as an example-based super-resolution [14] technique. Example-based methods leverage the high resolution information extracted from a HR image—an MPRAGE, for example—in conjunction with a LR input image—corresponding FLAIR image—to generate a SR version of the LR image. We used HR $(1 \times 1 \times 1\,mm^3)$ MPRAGE and FLAIR data, and downsampled the HR FLAIR to create a LR $(1 \times 1 \times 4\,mm^3)$ FLAIR. The atlas data included an HR MPRAGE + LR FLAIR and we trained SyCRAFT to predict the HR FLAIR. Given a test HR MPRAGE and LR FLAIR, we applied SyCRAFT to synthesize a SR FLAIR. We ran the LesionTOADS [16] segmentation algorithm on three scenarios for each subject: (a) HR MPRAGE + LR FLAIR (b) HR MPRAGE + SR FLAIR (c) HR MPRAGE + HR FLAIR. The last case acting as the ground truth for how the segmentation algorithm should behave on best case data. We aim to show that the tissue segmentation using SR FLAIR is closer to that achieved using HR FLAIR, than using LR

FLAIR. Figure 5(d) shows the super-resolution results, the LR FLAIR image is shown in Fig. 5(b), and the HR FLAIR image in Fig. 5(c). The corresponding LesionTOADS segmentations are shown in Figs. 5(e, f, g), respectively. The lesion boundaries as well as the cortex is overestimated when a LR FLAIR is used. Shown in Fig. 6 are the lesion volumes on 13 subjects for each of the three scenarios.
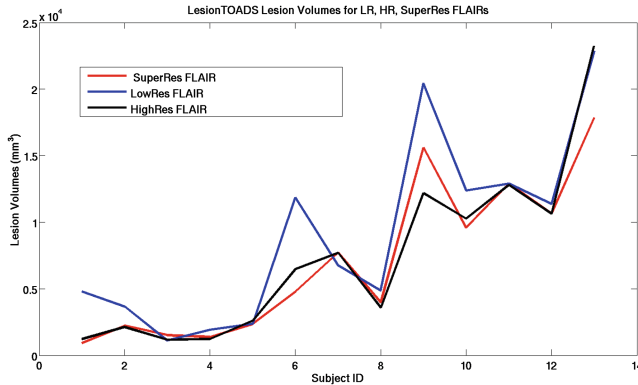

Fig. 6. Shown are the lesion volumes acquired by LesionTOADS on HR FLAIR+HR MPRAGE (black), LR FLAIR+HR MPRAGE (blue), and SR FLAIR+HR MPRAGE (red). Note that the black plot is closer to the red plot than the blue plot for all but one of the subjects (Colour figure online).

4   Conclusion

We have described an image synthesis framework, SyCRAFT, as a learning and inference problem on a Gaussian CRF. The parameters of the Gaussian CRF are built from parameters stored at the leaves of a single regression tree. Parameter learning is done by maximizing a pseudo-likelihood objective function. Our approach is extremely flexible in terms of features it can use to create the initial regression tree. It is also general enough to add larger neighborhoods and long-range relationships among voxels. Adding more neighbors leads to additional parameters, but these can be stored in the same initial tree and we do not need to create any more trees. Our approach is also computationally efficient, training from millions of samples in 20–30 min, and inference taking less than five minutes. We compared SyCRAFT to competitive image synthesis algorithms and showed that the image quality is superior. We also demonstrated practical benefits of using our algorithm to synthesize FLAIRs and validated the synthesis by showing tissue segmentation equivalent to that obtained using real FLAIRs. This shows that our image synthesis algorithm can be used in realistic scenarios, where imaging data is missing and needs to be replaced by a feasible alternative. Finally we also applied our algorithm to enhance the resolution of low resolution FLAIRs and showed improved tissue segmentation as a result.

References

1. Bland, J.M., Altman, D.G.: Statistical Methods For Assessing Agreement Between Two Methods Of Clinical Measurement. The Lancet **327**(8476), 307–310 (1986)
2. Breiman, L., Friedman, J.H., Olshen, R.A., Stone, C.J.: Classification and Regression Trees. Wadsworth Publishing Company, U.S.A (1984)
3. Burgos, N., Cardoso, M.J., Modat, M., Pedemonte, S., Dickson, J., Barnes, A., Duncan, J.S., Atkinson, D., Arridge, S.R., Hutton, B.F., Ourselin, S.: Attenuation correction synthesis for hybrid PET-MR scanners. In: Mori, K., Sakuma, I., Sato, Y., Barillot, C., Navab, N. (eds.) MICCAI 2013, Part I. LNCS, vol. 8149, pp. 147–154. Springer, Heidelberg (2013)
4. Evans, A., Collins, D., Mills, S., Brown, E., Kelly, R., Peters, T.: 3D Statistical Neuroanatomical Models from 305 MRI volumes. In: Nuclear Science Symposium and Medical Imaging Conference, vol. 3, pp. 1813–1817 (1993)
5. Hertzmann, A., Jacobs, C.E., Oliver, N., Curless, B., Salesin, D.H.: Image Analogies. Proceedings of SIGGRAPH **2001**, 327–340 (2001)
6. Iglesias, J.E., Konukoglu, E., Zikic, D., Glocker, B., Van Leemput, K., Fischl, B.: Is synthesizing MRI contrast useful for inter-modality analysis? In: Mori, K., Sakuma, I., Sato, Y., Barillot, C., Navab, N. (eds.) MICCAI 2013, Part I. LNCS, vol. 8149, pp. 631–638. Springer, Heidelberg (2013)
7. Jancsary, J., Nowozin, S., Sharp, T., Rother, C.: Regression Tree Fields; An Efficient, Non-parametric Approach to Image Labeling Problems. In: CVPR, 2376–2383 (2012)
8. Jog, A., Roy, S., Carass, A., Prince, J.L.: Magnetic Resonance Image Synthesis through Patch Regression. In: 10th International Symposium on Biomedical Imaging (ISBI 2013), pp. 350–353 (2013)
9. Kumar, S., Hebert, M.: Discriminative random fields. Int. J. Comput. Vision **68**(2), 179–201 (2006)
10. Lafferty, J.D., McCallum, A., Pereira, F.C.N.: Conditional random fields: probabilistic models for segmenting and labeling sequence data. In: ICML 2001, pp. 282–289 (2001)
11. Landman, B.A., Huang, A.J., Gifford, A., Vikram, D.S., Lim, I.A.L., Farrell, J.A.D., Bogovic, J.A., Hua, J., Chen, M., Jarso, S., Smith, S.A., Joel, S., Mori, S., Pekar, J.J., Barker, P.B., Prince, J.L., van Zijl, P.: Multi-parametric neuroimaging reproducibility: a 3-T resource study. NeuroImage **54**(4), 2854–2866 (2011)
12. Miller, M.I., Christensen, G.E., Amit, Y., Grenander, U.: Mathematical textbook of deformable neuroanatomies. Proc. Natl. Acad. Sci. **90**(24), 11944–11948 (1993)
13. Nyúl, L.G., Udupa, J.K., Zhang, X.: New variants of a method of MRI scale standardization. IEEE Trans. Med. Imag. **19**(2), 143–150 (2000)
14. 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)
15. Roy, S., Carass, A., Prince, J.L.: Magnetic resonance image example based contrast synthesis. IEEE Trans. Med. Imag. **32**(12), 2348–2363 (2013)
16. Shiee, N., Bazin, P.L., Ozturk, A., Reich, D.S., Calabresi, P.A., Pham, D.L.: A topology-preserving approach to the segmentation of brain images with multiple sclerosis lesions. NeuroImage **49**(2), 1524–1535 (2010)
17. Tappen, M., Liu, C., Adelson, E., Freeman, W.: Learning gaussian conditional random fields for low-level vision. In: CVPR, pp. 1–8 (2007)
18. Wainwright, M.J., Jordan, M.I.: Graphical models, exponential families, and variational inference. Found. Trends Mach. Learn. **1**(1–2), 1–305 (2008)

19. Wang, Z., Bovik, A.C.: A universal image quality index. IEEE Signal Proc. Letters **9**(3), 81–84 (2002)
20. Wang, Z., Bovik, A.C., Sheikh, H.R., Member, S., Simoncelli, E.P.: Image quality assessment: from error visibility to structural similarity. IEEE Trans. Image Proc. **13**, 600–612 (2004)