A Hierarchical Ensemble Classifier for Multilabel Segmentation of Fat-Water MR Images

Faezeh Fallah and Bin Yang
Institute of Signal Processing
and System Theory, University of Stuttgart
Email: {faezeh.fallah, bin.yang}@iss.uni-stuttgart.de

Sven S. Walter and Fabian Bamberg
Department of Diagnostic and Interventional
Radiology, University Clinic of Tübingen
Email: sven.walter@med.uni-tuebingen.de

Abstract—Automatic segmentation of organs on fat-water magnetic resonance (MR) images not only enables an analysis of their morphological characteristics but also their tissues pathogenesis demonstrated by their fat fraction ratios. So far, only a few methods have been designed based on these images and all proposed segmentation algorithms have only addressed one organ at a time. In this paper, we propose a hierarchical deformation-/registration-free algorithm for multilabel segmentation of fat-water MR images without need to prior localizations or geometry estimations. This method involved a hierarchical random forest classifier and a hierarchical conditional random field (CRF) encoding a multi-resolution image pyramid. This pyramid was formed by extracting multiscale local and contextual features from image patches at different resolutions. The classifier used penalized multivariate linear discriminants and SMOTEBagging to mitigate limited and imbalanced training data. The CRF refined the segmentations with regard to the spatial and hierarchical consistencies of the labels by using layer-specific significant features identified over the trained random forest classifier. Also, we incorporated resolution-specific hyperparameters to handle variable numbers or class mixtures of the image patches over hierarchical structures. This method was trained and evaluated for segmenting 10 thoracic and 5 lumbar VBs and IVDs on 30 training and 30 test volumetric fat-water (2 channel) MR images. Objective evaluations revealed its comparable accuracy to the state-of-the-art while demanding less computational burden.

I. INTRODUCTION

Noncontrast enhanced magnetic resonance (MR) imaging provides a noninvasive means for automated image-based analysis of the morphological and functional properties of tissues. Manual segmentations are tedious, costly, and error-prone while well-designed automatic algorithms enable robust reproducible segmentations.

Automatic simultaneous segmentation of multiple organs on MR images could enable image-based analyzes while eliminating the burden of one-by-one segmentations. These simultaneous segmentations become more important when multiple organs or organs of disconnected parts are to be analyzed on large cohort images.

Most of the previous methods for automatically segmenting organs on MR images have relied on T1- or T2-weighted MR images [1]. Chemical-shift-encoded (Dixon) MR imaging provides fat and water images of high contrasts that enable a simultaneous assessment of tissue morphologies and its tissue pathogenesis reflected by fat fraction ratios [2]. Additionally, most of the state-of-the-art methods for segmenting organs have derived and initialized a local [1], [3] or a global/local [1], [2] model on shapes, geometries, pairwise geometric constraints,

or intensities of the addressed tissues to steer image registrations/deformations towards organs boundaries. These steering-based approaches were to reduce the computational burden of the registrations/deformations. However, they demanded a localization or geometry estimation prior to the segmentations and implied a sequential (one-by-one) localization and segmentation if multiple tissues or organs of disconnected parts were addressed. Besides the computational burden, the sequential localizations/segmentations suffered from ambiguities if the addressed tissues had a repetitive pattern. These ambiguities were addressed by spatial probabilistic maps obtained from trained CNN models [1], parzen windows applied to the training images, or hidden Markov models of distances, relative shapes or pose differences of neighboring tissues [1], [3]. Having significant shape, geometry, and intensity variations in the training images could enhance the generalizability of the derived models for segmentations. However, they could reduce the specificity of above methods in resolving the localization ambiguities.

In this paper, we propose a hierarchical deformation-/registration-free method for automatically segmenting multiple organs on fat-water MR images without need to prior localization or geometry estimation. Due to the repetitive pattern and the disconnected form of vertebral bodies (VBs) and intervertebral discs (IVDs), the proposed method was trained and evaluated for segmenting 10 thoracic (T3–T12) and 5 lumbar (L1–L5) human VBs and their IVDs. These were done using 30 volumetric training and 30 volumetric test fat-water MR images acquired from 37 men and 23 women of 42 ± 18 years age, 165 ± 15 cm height, and 33 ± 11 kg/m^2 body mass indices, using a two-point 3D Dixon-VIBE sequence on a 3T whole body clinical MR scanner.

Besides simultaneous multilabel segmentation, the proposed method is quite generic and applicable to any organs/tissues whose segmentations are challenged by anatomical complexity or ambiguities of automatic localizations.

II. MATERIALS AND METHODS

A. Generation of the Reference Masks (Ground Truths)

Volumetric reference masks of the addressed VBs and IVDs were manually segmented on the Dixon MR images using the tools of the Medical Imaging Interaction Toolkit (MITK) of release 2015.05 [4]. This way, each image voxel was assigned a label from $L = \{VBs, IVDs, BG\}$, where BG denoted background. These labels formed a map $\mathcal{L}^{(1)} : \Omega \rightarrow L$, with $\Omega \subset R^3$ being the domain of all the processed fat-water images. For each volunteer, these segmentations took around 6
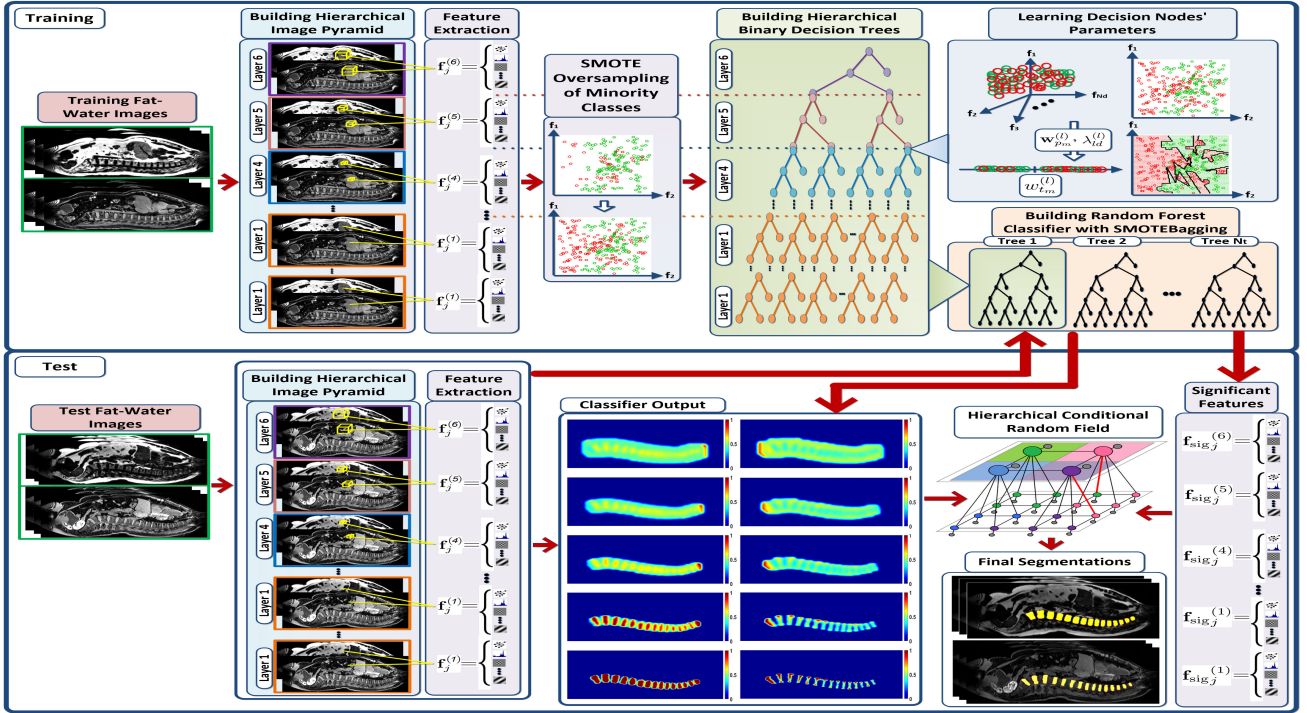
Fig. 1. Flow chart of the proposed framework for automatically segmenting multiple organs on fat-water MR images.

hours. These segmentations, especially on the boundaries, were subjective and prone to errors caused by image artifacts. Due to lack of ground truths for in-vivo images, they were considered to be the ground truths labels of automatic segmentations.

B. Preprocessing of MR Images

To reduce computations, images intensities were linearly normalized to the range [0, 128]. Also, according to [5], intensity nonuniformities were estimated on each volumetric fat image and were reduced on it and its corresponding water image. The preprocessed images and their ground truths were equally divided into a *training* and a *test* subset.

C. Framework of Automatic Segmentation

The proposed segmentation framework is shown in Fig. 1. It consisted in separate steps for the training and the test and included a hierarchical random forest classifier and a hierarchical conditional random field (CRF) based on a multiresolution image pyramid. During the training, multiresolution training data were used to optimize the random forest classifier. During the test, multiresolution test data were used to evaluate the performance of the segmentation method. The multiresolution training data were formed from the reference labels and features of *all* the training fat-water images. The multiresolution test data were formed by extracting same features from *each* test fat-water image.

D. Multiresolution Image Pyramid

The multiresolution image pyramid consisted of 6 resolution layers. In its rth layer, cubic image patches $\{q_j^{(r)} \in \mathcal{Q}^{(r)}\}$ of $2^{3(r-1)}$ voxels formed a sampling set $\mathcal{Q}^{(r)} = \bigcup_j q_j^{(r)}$ of the image domain $\Omega \subset R^3$. The finest image patches at $r = 1$ were the image voxels. That is, $\mathcal{Q}^{(1)} = \Omega$.

The sets $\{\mathcal{Q}^{(r)}\}_{r=1}^6$ were generated from coarse to fine by uniformly dividing each patch $q_j^{(r)} \in \mathcal{Q}^{(r)}$ into 8 disjoint patches via a map $\mathcal{M}^{(r)} : \mathcal{Q}^{(r)} \rightarrow \mathcal{Q}^{(r-1)}$. That is,

$\mathcal{M}^{(r)}(q_j^{(r)}) = \{q_i^{(r-1)}\} \subset \mathcal{Q}^{(r-1)}$ with $q_j^{(r)} = \bigcup_i q_i^{(r-1)}$. In $r = 6$, patches overlaps were $100 \times (1 - (2^1/2^5))\%$ in all directions. This way, larger patches were sampled denser.

The rth layer of the image pyramid was encoded in the rth layer of the hierarchical decision trees. To this end, $\mathcal{Q}^{(r)}$ and $\mathcal{M}^{(r)}$ were used to spatially decompose the received samples of the rth trees layer after classifying them.

E. Feature Extraction

A feature vector $\mathbf{f}_j^{(r)}$, consisting of local and contextual intra-channel and inter-channel features, was extracted from every patch $q_j^{(r)} \in \mathcal{Q}^{(r)}$ of every fat-water image. It contained 158 features for patches with more than one voxel and 67 features for single-voxel patches. The local intra-channel features were extracted from every fat or water patch and included median of intensities, average gradient magnitude, average gradient orientation, and 42 angle-invariant 3D Haralick features [6].

By calculating mean and maximum absolute feature differences between a fat and its corresponding water patch, between a fat (water) patch and its 26-connected fat (water) patches, and between a water patch and its 26-connected fat patches, the local inter-channel features, the contextual intra-channel features, and the contextual inter-channel features were obtained, respectively. Last features measured the distribution of fatty tissues around each lean tissues.

The optimality of the 26-connected neighborhood was shown previously [7]. To avoid instabilities of the penalized linear discriminants, features were normalized to zero mean and unit variance [8]. Also, use of mean and maximum features differences, made the above features rotation-invariant.

F. Training Data and Its Balancing

The multiresolution training data $\mathcal{T} = \{\mathcal{T}^{(r)}\}_{r=1}^6$ with $\mathcal{T}^{(r)} = \{t_j^{(r)} = (q_j^{(r)}, h_j^{(r)}, l_j^{(r)}, \mathbf{f}_j^{(r)})\}$ associated each patch $q_j^{(r)} \in \mathcal{Q}^{(r)}$ to its label histogram $h_j^{(r)} \in H$, its ground truth label

$l_j^{(r)} \in L$, and its feature vector $\mathbf{f}_j^{(r)}$. The label histogram $h_j^{(r)} \in H$ was used to extend the voxel-wise ground truths $\mathcal{L}^{(1)}$ to the patch-wise ground truths $\mathcal{L}^{(r)} : \mathcal{Q}^{(r)} \to L$ for labels $\{l_j^{l_j^{(r)}}\}_{r=1}^6$.

To avoid segmentation biases towards dominant (background) class, training samples of VBs and IVDs were oversampled by a multilabel Synthetic Minority Over-sampling Technique (SMOTE) [9]. We modified this approach by not only interpolating the feature vectors but also the label histograms of nearby samples of the same class. Also, we assigned dummy patches to the synthetic samples by updating $\{\mathcal{Q}^{(r)}\}_{r=1}^6$ and $\{\mathcal{M}^{(r)}\}_{r=1}^6$ to $\{\mathcal{Q}'^{(r)}\}_{r=1}^6$ and $\{\mathcal{M}'^{(r)}\}_{r=1}^6$. This allowed spatial decompositions of the synthetic patches as well as the original ones. Then according to the SMOTEBagging [9], for each decision tree, a balanced data $\mathcal{T}'_b = \{\mathcal{T}'^{(r)}_b\}_{r=1}^6$ with $\mathcal{T}'^{(r)}_b = \{t_j'^{(r)} = (q_j'^{(r)}, h_j'^{(r)}, l_j'^{(r)}, \mathbf{f}_j'^{(r)})\}$ were generated from $\mathcal{T} = \{\mathcal{T}'^{(r)}\}_{r=1}^6$. $b \in \{0.1, 0.2, ..., 1\}$ was a rate factor and every 10 trees processed balanced data of the same b [9].

G. Building the Hierarchical Binary Decision Trees

Each decision tree was recursively grown from coarse to fine by feeding the coarsest training data $\mathcal{T}'^{(6)}_b$ to its root node. A binary decision node $m^{(r)}$ at the rth tree layer, split its received data $\mathcal{T}'^{(r)}_m \subset \mathcal{T}'^{(r)}_b$ into its left and right child nodes by a penalized multivariate linear discriminant with parameter set $\mathbf{c}_m^{(r)} = (\mathbf{c}_{p_m}^{(r)}, c_{t_m}^{(r)}); c_{t_m}^{(r)} \in \{-1, -0.75, ..., +1\}$ was a threshold and $\mathbf{c}_{p_m}^{(r)} \in R^{N_d}$ were projecting coefficients given by

$$\mathbf{c}_{p_m}^{(r)} = \arg\min_{\hat{\mathbf{c}}_{p_m}^{(r)}} \frac{1}{|\mathcal{T}'^{(r)}_m|} \left(\left(\sum_{j=1}^{|\mathcal{T}'^{(r)}_m|} [a(\mathbf{f}''^{(r)}_j) - (\hat{\mathbf{c}}_{p_m}^{(r)T} \cdot \mathbf{f}''^{(r)}_j)]^2 \right) + \lambda_{ld}^{(r)} \sum_{i=1}^{N_d} |\hat{c}_{p_{m_i}}^{(r)}|^2 \right), \quad (1)$$

where $\mathbf{f}''^{(r)}_j \subset \mathbf{f}'^{(r)}_j$ was a vector of $N_d \approx \sqrt{\dim(\mathbf{f}'^{(r)}_j)}$ randomly sampled features at $m^{(r)}$; $-1 \leq \hat{c}_{p_{m_i}}^{(r)} \leq +1$ was the ith element of $\hat{\mathbf{c}}_{p_m}^{(r)}$ and $a : R^{N_d} \to \{-1, +1\}$ fulfilled the constraint $\sum_{j=1}^{|\mathcal{T}'^{(r)}_m|} |a(\mathbf{f}''^{(r)}_j)|^2 / |\mathcal{T}'^{(r)}_m| = 1$ required by casting the classification into a regression [8].

The above minimization was done by an adaptive iteratively re-weighted Penalized Least Squares (airPLS) algorithm [10]. The optimum set $\mathbf{c}_m^{*(r)} = (\mathbf{c}_{p_m}^{*(r)}, c_{t_m}^{*(r)})$ was the minimizer of the Gini index of the split at $m^{(r)}$:

$$\mathbf{c}_m^{*(r)} = \arg\min_{\forall (\mathbf{c}_{p_m}^{(r)}, c_{t_m}^{(r)})} GI_{\text{split}}(m^{(r)}). \quad (2)$$

Indices of $\mathbf{f}''^{(r)}_j \subset \mathbf{f}'^{(r)}_j$, labels distribution $l_{\mathcal{T}'^{(r)}_m}(l) = |\{t_j'^{(r)} \in \mathcal{T}'^{(r)}_m | l_j'^{(r)} = l\}|_{l \in L}$, empirical class posterior probabilities $\{P_m^{(r)}(l) = l_{\mathcal{T}'^{(r)}_m}(l) / (\sum_{l' \in L} l_{\mathcal{T}'^{(r)}_m}[l'])\}_{l \in L}$, and the optimum parameters $\mathbf{c}_m^{*(r)}$ were saved at each decision node $m^{(r)}$ to be used in the test phase.

The gini impurity of $\mathcal{T}'^{(r)}_m$ was $GI(\mathcal{T}'^{(r)}_m) = 1 - \sum_{l \in L} (P_m^{(r)}(l))^2$ and the impurity decrease after the split at $m^{(r)}$ was $\Delta GI_{\text{split}}(m^{(r)}) = GI(\mathcal{T}'^{(r)}_m) - GI_{\text{split}}(m^{(r)})$.

In $r = 1$, the decision nodes only processed single-voxel training data without further decomposition. A leaf node $m^{(1)}$

was made when its received data had $GI(\mathcal{T}'^{(1)}_m) \leq 10^{-5}$ or a preoptimized maximum tree depth D_t^* was reached.

H. Layer-wise Feature Selection

To reduce computations on the CRF, only most discriminant features, identified over the trained random forest classifier, were used to determine optimum labels. To this end, we calculated following weighted sum of coefficients over all the decision nodes of the rth layer of every tree $t \in N_t$,

$$\mathbf{c}_t^{(r)} = \frac{\sum_{m^{(r)} \in t} \Delta GI_{\text{split}}(m^{(r)}) \cdot \mathbf{c}'^{*(r)}_{p_m}}{\sum_{m^{(r)} \in t} \Delta GI_{\text{split}}(m^{(r)})}, \quad (3)$$

where $\mathbf{c}'^{*(r)}_{p_m}$ of $\dim(\mathbf{f}'^{(r)}_j)$ size was a vector obtained from taking absolute values of the elements of $\mathbf{c}_{p_m}^{*(r)}$ and assigning zero projecting coefficients to unprocessed features at $m^{(r)}$.

Then we computed the average $\overline{\mathbf{c}}^{(r)} = \frac{1}{N_t} \sum_{t \in N_t} \mathbf{c}_t^{(r)}$ and interpreted its fth element, $0 \leq \overline{c}_f^{(r)} \leq +1$, as the significance level of the fth element of $\mathbf{f}'^{(r)}_j$ or $\mathbf{f}_j^{(r)}$. Features with $\overline{c}_f^{(r)} \geq 0.25$ formed the vector of significant features $\mathbf{f}_{\text{sig}\,j}^{(r)}$ for every image patch at the rth layer.

I. Automatic Segmentation of the Test Data

The multiresolution test data $\mathcal{D} = \{\mathcal{D}^{(r)}\}_{r=1}^6$ with $\mathcal{D}^{(r)} = \{d_j^{(r)} = (q_j^{(r)}, \mathbf{f}_j^{(r)})\}$ were generated from *each* test fat-water image. The coarsest test data $\mathcal{D}^{(6)}$ were fed to the root node of the trained decision trees to be processed from coarse to fine until leaf nodes were reached. Every test sample $d_j^{(r)} \in \mathcal{D}_m^{(r)} \subset \mathcal{D}^{(r)}$ received by $m^{(r)}$, was affected the posterior probabilities $\{P_j^{(r)}(\hat{l}_j^{(r)} = l | \mathbf{f}''^{(r)}_j) = P_m^{(r)}(l)\}_{l \in L}$ for its label $\hat{l}_j^{(r)} \in L$. The labels distribution $l_{\mathcal{T}'^{(r)}_m}(l)_{l \in L}$, saved at $m^{(r)}$, determined an initial estimate of the label, $\hat{l}_{j_0}^{(r)}$, for $d_j^{(r)} \in \mathcal{D}_m^{(r)}$ as

$$\hat{l}_{j_0}^{(r)} = \arg\max_{l \in L} l_{\mathcal{T}'^{(r)}_m}(l). \quad (4)$$

The test sample $d_j^{(r)} \in \mathcal{D}_m^{(r)}$ was then split by the optimized discriminant at $m^{(r)}$ and decomposed into a finer resolution via the map $\mathcal{M}^{(r)}$ to be processed by the $(r-1)^{\text{th}}$ tree's layer. Each $d_j^{(r)} \in \mathcal{D}^{(r)}$ could be processed by $N_t' \leq N_t$ trees. Thus the average of its affected probabilities, $\{\overline{P}_j^{(r)}(l)\}_{l \in L}$, were considered for determining its label $\hat{l}_j^{(r)} \in L$:

$$\hat{l}_j^{(r)} = \hat{l}_j^{\text{pos}(r)} = \arg\max_{l \in L} \overline{P}_j^{(r)}(l). \quad (5)$$

The estimates $\hat{\mathbf{l}}^{\text{pos}} = \{\hat{l}_j^{\text{pos}(r)} \in L\}_{r=1}^6$ were regardless of spatial (neighborhood) or hierarchical relationships between the patches. To consider these, $\{\overline{P}_j^{(r)}(l), l \in L\}_{r=1}^6$ were fused and regularized by a hierarchical CRF defined on an undirected graph $\mathcal{G} = (\mathcal{V}, \mathcal{E}_a \bigcup \mathcal{E}_p)$ with vertices $\mathcal{V}$ and edges $\mathcal{E}_a \bigcup \mathcal{E}_p$ [11]. The graph was based on the multiresolution image pyramid with $\mathcal{E}_a = \{\mathcal{E}_a^{(r)}\}_{r=1}^6$, $\mathcal{E}_a^{(r)} = \{(i,j) | q_i^{(r)} \in \mathcal{Q}^{(r)}$ is *adjacent* to $q_j^{(r)} \in \mathcal{Q}^{(r)}\}$, $\mathcal{E}_p = \{\mathcal{E}_p^{(r)}\}_{r=2}^6$, and $\mathcal{E}_p^{(r)} = \{(i,j) | q_i^{(r)} \in \mathcal{Q}^{(r)}$ is *parent* of $q_j^{(r-1)} \in \mathcal{Q}^{(r-1)}\}$.

Over this graph, Eq. 6 was minimized by an iterative primal-dual algorithm [12]; $\hat{\mathbf{l}} = \{\hat{l}_j^{(r)} \in L\}_{r=1}^6$ were the estimated labels; $\delta(.)$ was the Kronecker function; $\|.\|_1$ was the l_1 norm, and $\mathbf{f}_{\text{sig}\,j}^{(r)}$ were the layer-wise significant features. The spatial (intra-layer) and the hierarchical (inter-layer) consistencies of

$$E(\hat{\mathbf{l}}) = -\Big(\sum_{r=1}^{6}\sum_{j=1}^{|\mathcal{D}^{(r)}|}\log(\{\overline{P}_j^{(r)}(r)\}_{l\in L})\Big) + \Big(\sum_{r=1}^{6}\sum_{\forall(i,j)\in\mathcal{E}_a^{(r)}}\lambda_s^{(r)}\cdot\exp\big(-\|\mathbf{f}_{\mathrm{sig}\,i}^{(r)}-\mathbf{f}_{\mathrm{sig}\,j}^{(r)}\|_1/\dim(\mathbf{f}_{\mathrm{sig}\,i}^{(r)})\big)\cdot(1-\delta(\hat{l}_i^{(r)},\hat{l}_j^{(r)}))\Big)$$

$$+ \Big(\sum_{r=2}^{6}\sum_{\forall(i,j)\in\mathcal{E}_p^{(r)}}\lambda_h\cdot\exp\big(-\|\mathbf{f}_{\mathrm{sig}\,i}^{(r)}-\mathbf{f}_{\mathrm{sig}\,j}^{(r-1)}\|_1/\dim(\mathbf{f}_{\mathrm{sig}\,i}^{(r)})\big)\cdot(1-\delta(\hat{l}_i^{(r)},\hat{l}_j^{(r-1)}))\Big), \quad (6)$$

TABLE I
QUANTITATIVE METRICS COMPARING THE AUTOMATICALLY (AU) AND THE MANUALLY (GT) SEGMENTED OBJECTS

Name	Definition
Dice (%)	$\frac{2|AU\cap GT|}{|AU|+|GT|}$
MSSD (mm)	$\frac{\sum_{\mathbf{a}\in\delta(AU)}\min_{\mathbf{g}\in\delta(GT)}d(\mathbf{a},\mathbf{g}) + \sum_{\mathbf{g}\in\delta(GT)}\min_{\mathbf{a}\in\delta(AU)}d(\mathbf{a},\mathbf{g})}{\mathrm{card}(AU\cup GT)}$
HSD (mm)	$\max\big(\max_{\mathbf{g}\in\delta(GT)}\min_{\mathbf{a}\in\delta(AU)}d(\mathbf{a},\mathbf{g}),\ \max_{\mathbf{a}\in\delta(AU)}\min_{\mathbf{g}\in\delta(GT)}d(\mathbf{a},\mathbf{g})\big)$

$d(\mathbf{a},\mathbf{g})$ was the Euclidean distance between $\mathbf{a}$ and $\mathbf{g}$; δ denoted the surface. Dice compared the segmented volumes. However, MSSD and HSD compared the segmented surfaces.

TABLE II
THE OPTIMIZED HYPERPARAMETER VALUES

Hyperparameter	N_t^*	D_t^*	$\lambda_{ld}^{*(6)}$	$\lambda_{ld}^{*(5)}$	$\lambda_{ld}^{*(4)}$	$\lambda_{ld}^{*(3)}$	$\lambda_{ld}^{*(2)}$	$\lambda_{ld}^{*(1)}$
Value	80	15	10^1	10^1	10^0	10^{-2}	10^{-3}	10^{-1}

Hyperparameter	$\lambda_s^{*(6)}$	$\lambda_s^{*(5)}$	$\lambda_s^{*(4)}$	$\lambda_s^{*(3)}$	$\lambda_s^{*(2)}$	$\lambda_s^{*(1)}$	λ_h^*
Value	10^{-4}	10^{-3}	10^{-2}	10^{-2}	10^{-1}	10^1	10^{-1}

N_t: Number of the decision trees; D_t: Maximum depth of the trees; $\{\lambda_{ld}^{(r)}\}_{r=1}^6$: Regularization parameters of the linear discriminants; $\{\lambda_s^{(r)}\}_{r=1}^6$: Parameters for spatial regularization of the labels; λ_h: Parameter for hierarchical regularization of the labels.

labels were regularized by the second and the third term of Eq. 6, giving the optimum labels $\hat{\mathbf{l}}^{\mathrm{crf}} = \{\hat{l}_j^{\mathrm{crf}(r)}\in L\}_{r=1}^6$ as

$$\hat{\mathbf{l}}^{\mathrm{crf}} = \arg\min_{\hat{\mathbf{l}}} E(\hat{\mathbf{l}}), \quad (7)$$

To speed up this minimization, each $\hat{l}_j^{(r)}$ was initialized by $\hat{l}_{j_0}^{(r)}$ given by Eq. 4. In Eq. 6, by replacing $\hat{\mathbf{l}}$ with $\hat{\mathbf{l}}^{\mathrm{crf}}$ and merging the regularization terms in their corresponding probabilities, the refined probabilities $\{P_j^{\mathrm{crf}(r)}(l)\}_{l\in L}$, fulfilling

$$E_{\min} = E(\hat{\mathbf{l}}^{\mathrm{crf}}) = -\sum_{r=1}^{6}\sum_{j=1}^{|\mathcal{D}^{(r)}|}\log(\{P_j^{\mathrm{crf}(r)}(l)\}_{l\in L}), \quad (8)$$

were obtained. The voxel-wise labels ($\{\hat{l}_j^{\mathrm{pos}(1)}\}$ or $\{\hat{l}_j^{\mathrm{crf}(1)}\}$), determined the volumetric masks of the automatically segmented organs on the test fat-water image.

J. Evaluation

The proposed segmentation framework was objectively evaluated using the quantitative metrics of Dice coefficient (Dice), mean symmetric surface distance (MSSD), and Hausdorff distance (HSD) [1] described in Table I.

III. RESULTS

By paralleling trees' predictions on a quad-core CPU of 3.10 GHz frequency and 16 GB RAM, this method segmented 15 VBs and 15 IVDs of a test subject in 11 ± 2 min.

Table II shows the optimized hyperparameters of the proposed segmentation framework. Fig. 2 shows the voxel-wise probabilities and the segmented masks of the VBs and IVDs on mid-sagittal slices of 3 test fat-water images. Fig. 3 shows the box plots of the quantitative metrics, described

in Table I, for the automatically segmented volumes of the VBs and IVDs on 30 test fat-water images. In segmenting VBs on all the test fat-water images, this method achieved Dice coefficient (Dice) = $92.5\pm1.9\%$, mean symmetric surface distance (MSSD) = 0.65 ± 0.18 mm, and Hausdorff distance (HSD) = 3.98 ± 1.12 mm. For IVDs, it achieved Dice = $91.4\pm2.3\%$, MSSD = 1.18 ± 0.4 mm, HSD = 4.6 ± 0.9 mm. These results compared favorably to the state-of-the-art for segmenting VBs/IVDs on MR images [1], [3], [13].

In the erroneous segmentations of the present method, two major patterns were observed. One appeared at the boundaries and another was randomly distributed over the tissues. The former could be attributed to the limited training data and the latter to the use of an unsupervised CRF for refining the segmentations. Moreover, image artifacts or pathological tissue alterations could lead to either of these patterns.

IV. DISCUSSION AND CONCLUSION

In this paper, we proposed a registration-/deformation-free method for automatically segmenting multiple organs of disconnected parts on volumetric fat-water MR images without prior localizations or geometry estimations. The proposed method extended [11], [14], [15] by

- encoding a multiresolution image pyramid in each binary decision tree for having one classifier for all resolutions
- handling limited training data by using a penalized multivariate linear discriminant at every decision node
- handling different number of training data in different trees' layers by defining resolution-specific regularization parameters $\{\lambda_{ld}^{(r)}\}_{r=1}^6$ for the penalized discriminants
- denser sampling of coarser patches by defining a resolution-specific patch sampling
- feasible node optimization despite of large feature dimension by combining analytical and exhaustive searches
- mitigating class imbalances of the training data by using a modified SMOTEBagging [9] in the hierarchical classifier
- using a new set of multiscale local and contextual features for a fast localization and a simultaneous segmentation
- applying a layer-wise feature selection and using these features and the label distributions of the training data for speeding up the estimates on the CRF
- handling different class mixtures of image patches at different resolution layers by defining resolution-specific hyperparameters for the hierarchical CRF

Major bottlenecks of the present method were using limited training data, refining segmentations by an unsupervised CRF, and ignoring the effects of image artifacts or pathological tissue alterations on the extracted features. Also, the derived segmentation model was dependent on the modality, contrast, and quality of the used images and could be changed if these characteristics vary. Future works would be incorporating multimodal, multicenter, and pathological images for a higher clinical utility. Application of this method to multichannel
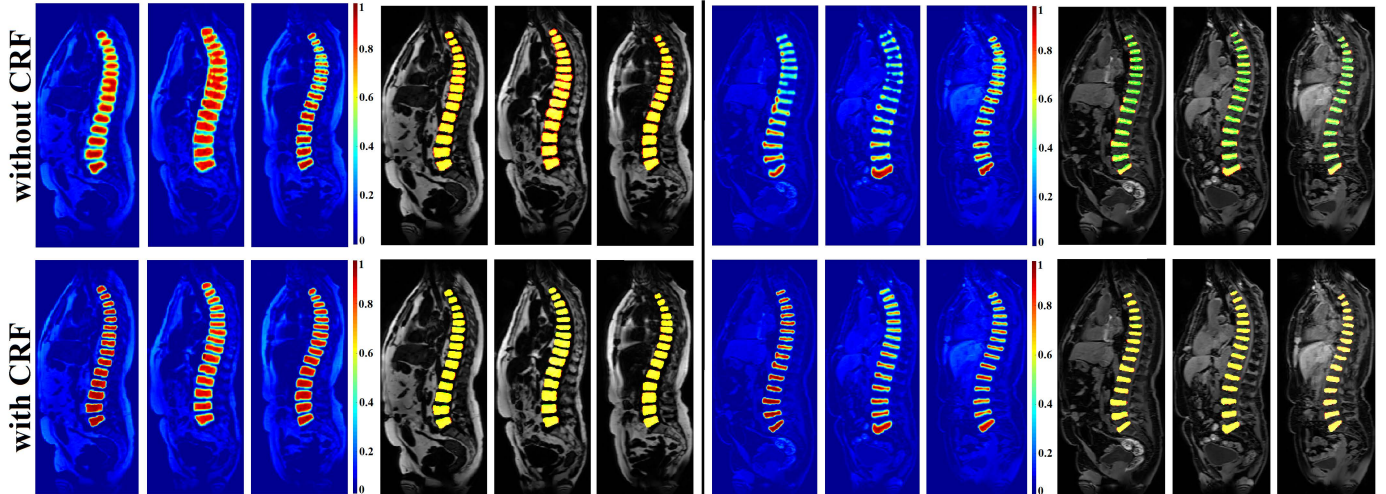
Fig. 2. **Left Parts:** Voxel-wise probabilities $\{\overline{P}_j^{(1)}(r)\}_{l\in\{VBs,IVDs\}}$ (without CRF) or $\{P_j^{\mathrm{crf}(1)}(r)\}_{l\in\{VBs,IVDs\}}$ (with CRF) of VBs and IVDs of 3 test fat-water images shown on their mid-sagittal slices. **Right Parts:** Automatically (red) and manually (green) segmented masks and their overlaps (yellow) for the VBs and IVDs of the same slices. The automatic masks were based on $\{\hat{t}_j^{\mathrm{pos}(1)}\}$ (without CRF) or $\{\hat{t}_j^{\mathrm{crf}(\bar{1})}\}$ (with CRF).
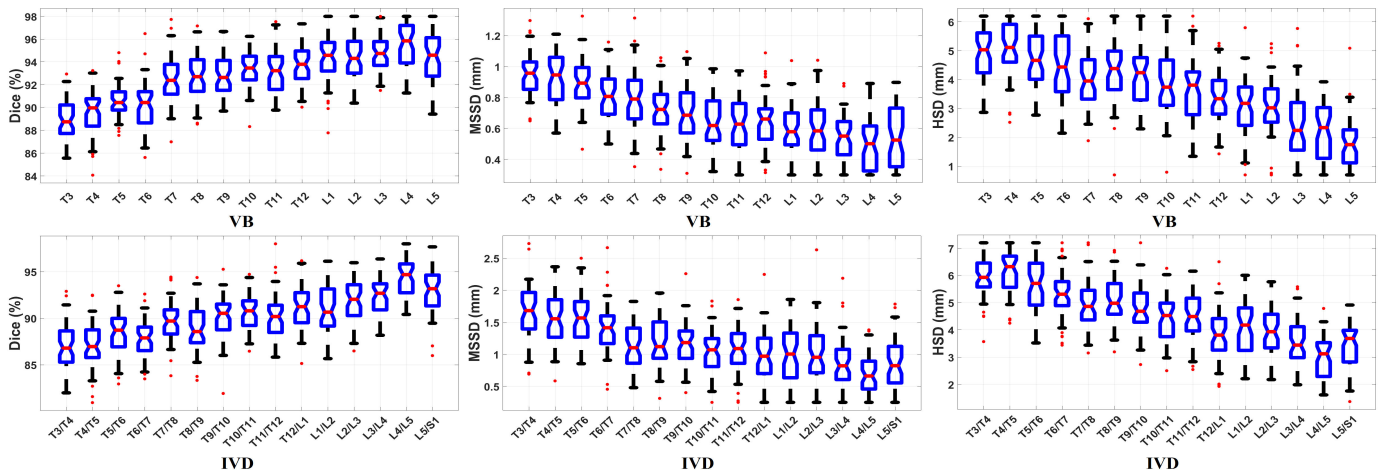

Fig. 3. Box plots of the quantitative metrics for the automatically segmented volumes of the VBs and IVDs on 30 test fat-water images.

medical images is straightforward.

REFERENCES

[1] G. Zheng, C. Chu, D. L. Belavý, B. Ibragimov, R. Korez, T. Vrtovec, H. Hutt, R. Everson, J. Meakin, I. L. Andrade, B. Glocker, H. Chen, Q. Dou, P.-A. Heng, C. Wang, D. Forsberg, A. Neubert, J. Fripp, M. Urschler, D. Stern, M. Wimmer, A. A. Novikov, H. Cheng, G. Armbrecht, D. Felsenberg, and S. Li, "Evaluation and comparison of 3D intervertebral disc localization and segmentation methods for 3D T2 MR data: A grand challenge," *Med. Image Anal.*, vol. 35, pp. 327–344, 2017.

[2] F. Fallah, B. Yang, and F. Bamberg, "Automatic atlas-guided constrained random walker algorithm for 3D segmentation of muscles on water magnetic resonance images," 2017, pp. 251–255.

[3] R. Korez, B. Ibragimov, B. Likar, F. Pernuš, and T. Vrtovec, "Intervertebral disc segmentation in MR images with 3D convolutional networks," vol. 1013306, 2017, pp. 1–10.

[4] I. Wolf, M. Vetter, I. Wegner, T. Böttger, M. Nolden, M. Schöbinger, M. Hastenteufel, T. Kunert, and H. P. Meinzer, "The medical imaging interaction toolkit," *Med. Image Anal.*, vol. 9, no. 6, pp. 594–604, 2005.

[5] F. Fallah, J. Machann, P. Martirosian, F. Bamberg, F. Schick, and B. Yang, "Comparison of T1-weighted 2D TSE, 3D SPGR, and two-point 3D Dixon MRI for automated segmentation of visceral adipose tissue at 3 Tesla," *Magn. Reson. Mater. Phy.*, vol. 30, no. 2, pp. 139–151, 2017.

[6] F. Han, H. Wang, G. Zhang, H. Han, B. Song, L. Li, W. Moore, H. Lu, H. Zhao, and Z. Liang, "Texture feature analysis for computer-aided diagnosis on pulmonary nodules," *J Digit Imaging*, vol. 28, no. 1, pp. 99–115, 2015.

[7] L. Grady and M.-P. Jolly, "Weights and topology: A study of the effects of graph construction on 3D image segmentation," 2008, pp. 153–161.

[8] B. H. Menze, B. M. Kelm, D. N. Splitthoff, U. Koethe, and F. A. Hamprecht, "On oblique random forests," 2011, pp. 453–469.

[9] S. Wang and X. Yao, "Diversity analysis on imbalanced data sets by using ensemble models," 2009, pp. 324–331.

[10] Z.-M. Zhang, S. Chen, and Y.-Z. Liang, "Baseline correction using adaptive iteratively reweighted penalized least squares," *Analyst*, vol. 135, pp. 1138–1146, 2010.

[11] V. Zografos, A. Valentinitsch, M. Rempfler, F. Tombari, and B. Menze, "Hierarchical multi-organ segmentation without registration in 3D abdominal CT images," 2016, pp. 37–46.

[12] N. Komodakis, G. Tziritas, and N. Paragios, "Performance vs computational efficiency for optimizing single and dynamic MRFs: Setting the state of the art with primal-dual strategies," *Comput. Vis. Image Underst.*, vol. 112, no. 1, pp. 14–29, 2008, http://www.csd.uoc.gr/~komod/FastPD/.

[13] Y. Cai, S. Osman, M. Sharma, M. Landis, and S. Li, "Multi-modality vertebra recognition in arbitrary views using 3D deformable hierarchical model," *IEEE Trans Med Imaging*, vol. 34, no. 8, pp. 1676–1693, 2015.

[14] C. Chu, D. L. Belavy, G. Armbrecht, M. Bansmann, D. Felsenberg, and G. Zheng, "Fully automatic localization and segmentation of 3D vertebral bodies from CT-MR images via a learning-based method," *PLoS ONE*, vol. 10, no. 11, pp. 1–22, 2015.

[15] E. Geremia, B. H. Menze, and N. Ayache, "Spatially adaptive random forest," 2013, pp. 1344–1347.