Localization and Segmentation of 3D Intervertebral Discs in MR Images by Data Driven Estimation

Cheng Chen, Daniel Belavy, Weimin Yu, Chengwen Chu, Gabriele Armbrecht,
Martin Bansmann, Dieter Felsenberg, and Guoyan Zheng*, *Member, IEEE*

***Abstract*—This paper addresses the problem of fully-automatic localization and segmentation of 3D intervertebral discs (IVDs) from MR images. Our method contains two steps, where we first localize the center of each IVD, and then segment IVDs by classifying image pixels around each disc center as foreground (disc) or background. The disc localization is done by estimating the image displacements from a set of randomly sampled 3D image patches to the disc center. The image displacements are estimated by jointly optimizing the training and test displacement values in a data-driven way, where we take into consideration both the training data and the geometric constraint on the test image. After the disc centers are localized, we segment the discs by classifying image pixels around disc centers as background or foreground. The classification is done in a similar data-driven approach as we used for localization, but in this segmentation case we are aiming to estimate the foreground/background probability of each pixel instead of the image displacements. In addition, an extra neighborhood smooth constraint is introduced to enforce the local smoothness of the label field. Our method is validated on 3D T2-weighted turbo spin echo MR images of 35 patients from two different studies. Experiments show that compared to state of the art, our method achieves better or comparable results. Specifically, we achieve for localization a mean error of 1.6–2.0 mm, and for segmentation a mean Dice metric of 85%–88% and a mean surface distance of 1.3–1.4 mm.**

***Index Terms*—Magnetic Resonance Imaging (MRI), spine, intervertebral discs, segmentation.**

I. INTRODUCTION

IN clinical practice, spine MRI is the preferred modality in diagnosis and treatment planning of various spinal pathologies such as disc herniation, slipped vertebra and so on. Accurate identification of intervertebral discs (IVD) on spine MR image is therefore very important for correct diagnosis and treatment planning [1], [2]. Traditionally, most quantitative studies on IVD degeneration have been done by manually localizing and segmenting the data [3], [4], which is tedious, time-consuming and error-prone. On the other hand, a fully-automatic system for IVD identification will significantly reduce the time of the diagnosis. An automatic system might also help reduce errors caused by subjective factors and improve the consistency of diagnosis standards. In this way, it can immediately benefit clinical applications and spinal biomechanics research.

In this paper, we are interested in the automatic *localization* and *segmentation* of IVDs from 3D spine MR images. Localization means identifying the location of each IVD center, and segmentation produces the binary labeling of the image into disc/non-disc regions (from which a 3D surface of the disc boundary can also be extracted if needed). We first conduct the localization, whose result is then used to reduce the size of the image region for segmentation.

Despite the research prevalence in brain, heart and bone images, we found relatively less study on spinal image with regard to vertebra or IVD localization [5]–[9] and segmentation [10]–[17]. In the work of Schmidt *et al.* [5], the IVDs were localized and labeled by a probabilistic model considering image intensity and geometric constraints. Corso *et al.* [6] enforced the inter-disc distance constraint to improve the label accuracy. Glocker *et al.* applied the Random Forest (RF) regression [7] and classification [8] to localize the vertebra body. In a recent work, Marjor *et al.* [9] detect the spinal canal, IVDs, T1 and T12 ribs and the uppermost sacral foramina of the sacrum by probabilistic boosting trees and Markov Random Field matching.

For vertebra or IVD segmentation, there exist methods based on Hough Transform [10], watershed algorithm [11], atlas registration [12], Adaboost and normalized-cut [13], graph cuts with geometric priors from neighboring discs [14] and template matching and statistic shape model [15]. In [15], Neubert *et al.* segment the IVD and vertebral body by analysing the image intensity profile. They first identify the 3D spine curve and localize the IVDs using Canny edge detector and the symmetry of image intensities. Then, the 3D mean shape model is placed on the locations, and iterative refinement is conducted by matching the image intensity profile of each mesh vertex. This method relies on the low level image information such as intensities and edges, which might be sensitive to noises and/or local anatomic deformation. In addition, the refinement based on intensity profile is local. In contrast, in our paper we use our data-driven regression and classification method which finds the global optimum. Recently, Law *et al.* [16] proposed a method based on anisotropic oriented flux detection, where they first track the position and orientation of the discs, and then the segmentation is

Manuscript received November 14, 2014; revised January 20, 2015; accepted January 22, 2015. Date of publication February 12, 2015; date of current version July 29, 2015. *Asterisk indicates corresponding author.*

C. Chen, W. Yu, and C. Chu are with the Institute of Surgical Technology and Biomechanics, University of Bern, CH-3014 Bern, Switzerland (e-mail: cheng.chen@istb.unibe.ch; weimin.yu@istb.unibe.ch; chengwen.chu@istb.unibe.ch).

D. Belavy, G. Armbrecht, and D. Felsenberg are with Charite University Medicine Berlin, 12203 Berlin, Germany.

M. Bansmann is with KardioMR Kln-Bonn, Krankenhaus Porz am Rhein, 51149 Köln, Germany.

*G. Zheng is with the Institute of Surgical Technology and Biomechanics, University of Bern, CH-3014 Bern, Switzerland (e-mail: guoyan.zheng@istb.unibe.ch).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TMI.2015.2403285

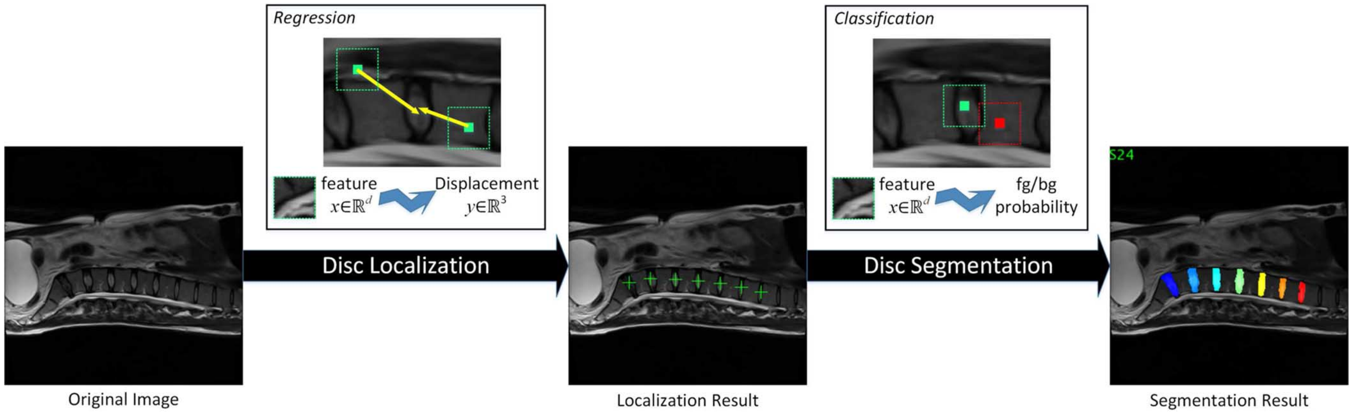


Fig. 1. Overview of our method which consists of the localization step followed by the segmentation step. The localization step is based on estimation of the image displacements of some image patches, while the segmentation step involves the classification of image pixels as foreground (inside disc) or background (outside disc).

conducted by minimizing an objective function using a level set based active contour model. Kelm *et al.* [17] propose an object detection method based on Marginal Space Learning [18]. The idea is that, instead of estimating the position, orientation and scale of the object at a time, they decompose the problem into three steps with three classifiers, where the first classifier is trained to estimate only the object position (marginal with regard to orientation and scale), the second classifier is trained to estimate both position and orientation, and the third classifier deals with all parameters (position, orientation and scale). In this way the efficiency is improved because the parameter searching space is reduced in each step. They use this strategy to localize the IVDs, and then the segmentation is done by case-adaptive graph cut [19]. Please note that most of the mentioned methods except [15] and [17] work on 2D sagittal images instead of 3D volumes. Also, many approaches (except [11], [13], [15] and [17]) require some manual user interaction such as segmentation of the first disc or at least localizing a point inside the disc.

In this paper, we propose a new fully-automatic method to localize and segment 3D IVDs from MR images. Fig. 1 gives an overview of our entire pipeline. Inspired by a recent development introduced in [20], where a data-driven approach is proposed to estimate image displacements and then to localize landmarks on 2D pelvis X-ray images, we formulate our localization problem into a similar framework of data-driven regression of image displacements. To estimate the disc center, we randomly sample a set of 3D image patches over the image, and for each patch we estimate its displacement to the unknown disc center by the data-driven regression on the patch appearance. Combining the predictions of all the sampled image patches, we can get an accurate estimation of the disc center. This process is done independently for each IVD. Finally, to exploit the inter-disc relations, we employ dynamic programming to post-process the localization results of each disc to resolve ambiguity caused by the repetitive pattern of IVDs.

Once the disc center localization is done, the segmentation of IVD is conducted by classification of each image pixel around the disc center as foreground (disc) or background. To this end, for each pixel we compute two probabilities of being foreground: the spatial probability based on the spatial relation of the pixel with regard to the associated disc center (which is encoded via Parzen windows), and the observation probability based on the visual feature of this pixel. The observation probability is calculated in a similar data-driven framework as we used for localization. The difference is that, instead of estimating the image displacements, now we estimate the foreground probability of the image pixels. In addition, due to the nature of segmentation, we introduce a new term in the objective function for segmentation, which enforces the neighborhood smoothness. The combination of the two measurements (spatial probability and appearance probability) gives the final probability of this pixel being inside the pixel disc, from which we derive the binary image segmentation.

We validated our method on T2-weighted MR spine images from a total of 35 patients, where the data come from two datasets (23 patients in dataset 1 and 12 patients in dataset 2). We conducted both leave-one-out study within each dataset as well as cross-dataset evaluations. We compared our performance to the state-of-the-art method based on Random Forest and found that our method achieved better or comparable results. Specifically, we achieved for localization a mean error of 1.6–2 mm, and for segmentation a mean Dice metric of 85%–88% and a mean surface distance of 1.3–1.4 mm.

In summary, we made three contributions in this paper.

1) We adapt the data-driven estimation framework of [20] which deals with landmark detection in 2D pelvis X-ray image into IVD localization in 3D MR spine image.
2) We extend the above mentioned data-driven approach into the segmentation domain, where we estimate the foreground probability instead of image displacement. We also introduce a new neighborhood smoothing term which is uniquely used for disc segmentation.
3) Combining our localization and segmentation together, we achieve a fully-automatic pipeline of IVD localization and segmentation, which is validated on 3D T2-weighted turbo spin echo MR images of 35 patients from two different studies.

We would like to point out that our data-driven regression/classification method can also be used in other tasks such as image segmentation and registration [26], [27].

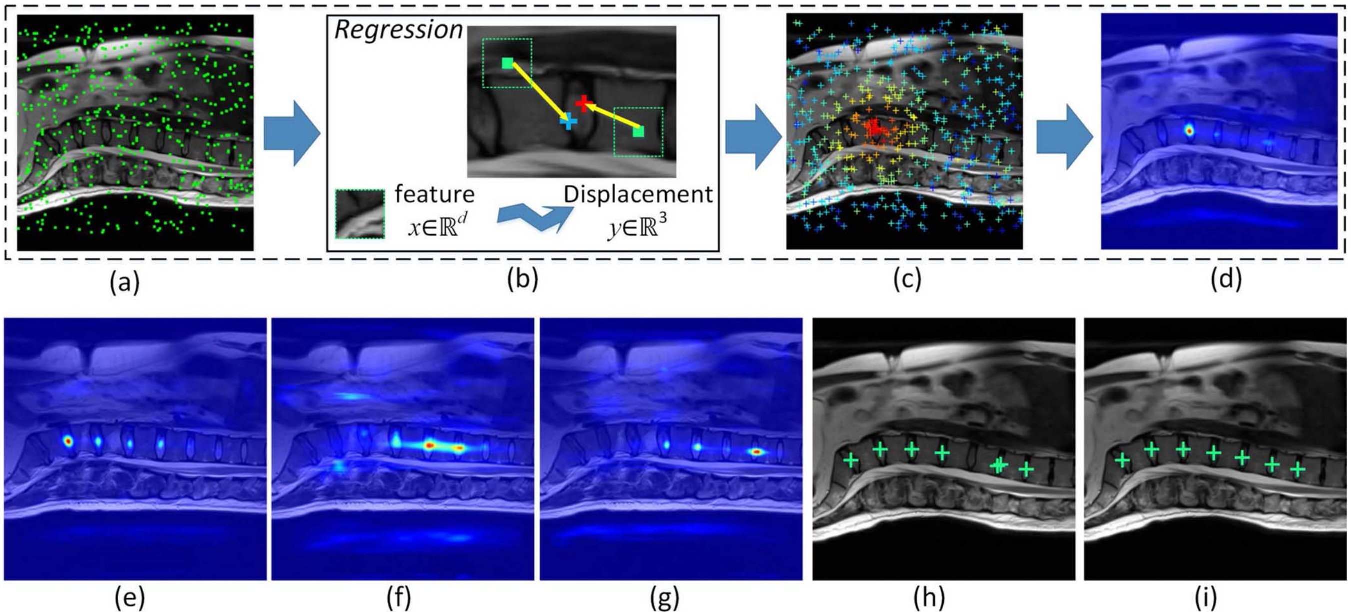


Fig. 2. Overview of the IVD localization method. Top row: the first step. Bottom row: the second step. Please see text in Section II.A for details.

The paper is organized as follows. Section II presents our disc center localization method. Section III presents our disc segmentation algorithm. Experiment results are shown in Section IV and we conclude the paper in Section V.

II. IVD LOCALIZATION BY DATA-DRIVEN IMAGE DISPLACEMENT REGRESSION

A. Overview

Fig. 2 gives an overview of our IVD localization method. Please note that for the simplicity of illustration, Fig. 2 is drawn on a 2D image slice, but our operation is on the 3D image. The localization is conducted in two steps. In the first step (the first row of Fig. 2), we try to estimate the location of each individual IVD separately. For each IVD, we randomly sample some locations in the image, as shown in Fig. 2(a). For each of these locations, as shown in Fig. 2(b), we calculate its visual feature, and based on the feature we estimate the image displacement from this location to the disc center. In this way, each sampled image location will make a vote about the disc center location, as shown in Fig. 2(c), where the dots represent the locations of votes and the colors encode the vote confidence. Each individual vote can be quite noisy, but when we aggregate them together, as in Fig. 2(d), we can approach a very accurate probability map of the disc center location. In this way, *for each IVD* we get a probability map. This step is explained in Section II.B. Please note that this process is repeated independently for each IVD, i.e., for each IVD we solve an optimization problem and get a probability map as in Section II.B.

After we get the probability maps of all IVDs, in the second step (the bottom row of Fig. 2), we exploit the inter-disc relation to regularize the results. This is because, the resultant probability maps can be multi-modal with potential ambiguities. For example, in Fig. 2(e-g) the probability maps of three discs are shown, where the ground-truth disc center locations are given by the orange arrows. We see that on the probability maps we have responses not only on the true location but also on some nearby discs. This is due to the repetitive appearance pattern of discs. If we simply take the mode of each probability map, sometimes we will get the wrong answer (e.g., in Fig. 2(h) a disc center location is wrong). We adopt a dynamic programming scheme to find the optimal disc center positions. Fig. 2(i) shows the final localization result. Section II.C explains this step in more details.

B. Localizing Individual IVDs by Estimating Image Displacement

The key component of the algorithm is the estimation of image displacements from the feature vector (Fig. 2(b)). The estimation is based on training data consisting of feature vectors with ground-truth displacements, which is a regression problem. There already exist methods to solve such a regression problem, such as Random Forest regression methods [21]–[24]. Recently, Chen *et al.* [20] proposed a new method to do this regression, where they formulate the problem into a joint optimization problem on the training and test displacements. The advantage of this method is that the displacements are not estimated independently (as in Random Forest methods) but are regularized by the geometric constraint.

Inspired by the method introduced in [20], in this paper we also follow the same framework but we make an improvement.

Notations: Suppose that N image locations are sampled on the training images, and let $\{x_i\}_{i=1\ldots N}$ denote the features calculated at these locations, where $x_i \in \mathbb{R}^d$. We denote $X = [x_1 \ldots x_N] \in \mathbb{R}^{d\times N}$. We use $\{y\}_{i=1\ldots N}$ to denote the image displacement vectors of the training locations, i.e., $y_i \in \mathbb{R}^3$, which are the displacement vectors from the ground-truth disc center to the sampled image locations in the training image. The training images are annotated, so that the ground-truth output values of training locations are known as $\{y^{\mathrm{GT}}{}_i\}_{i=1\ldots N}$, and we denote $Y^{\mathrm{GT}} = [y_1^{\mathrm{GT}} \ldots y_N^{\mathrm{GT}}]$.

Given a new image, we randomly sample N' locations at locations $\{c'\}_{i=1\dots N'}$, whose features are $\{x'\}_{i=1\dots N'}$, where $x'_i \in \mathbb{R}^d$. We denote $X' = [x'_1 \dots x'_{N'}]$. The task is to compute the corresponding image displacements $\{y'\}_{i=1\dots N'}$, which are the displacement vectors from the ground-truth disc center to the sampled image locations on the test image. We write $Y' = [y'_1 \cdots y'_{N'}]$.

Optimization Function: We solve for Y' by optimizing an objective function as defined in (1).

First, we construct a matrix $\tilde{Y} = [Y, Y']$ which is the composition of training and test image displacements. Although we want to compute Y', our objective function is defined on $\tilde{Y}$. In this way we can encode the relations between training and test data in a uniform way. After solving for the optimal $\tilde{Y}$, we simply take its right part as $Y' = \tilde{Y}Q$, with $Q = [\begin{smallmatrix}\mathbf{0}_{N\times N'}\\ \mathbf{I}_{N'\times N'}\end{smallmatrix}]$ as the matrix selecting the right part, where $\mathbf{0}_{N\times N'}$ is a zero matrix with all its entries being zero and $\mathbf{I}_{N'\times N'}$ is an identity matrix.

The objective function consists of three terms:

$$E(\tilde{Y}) = E_g(\tilde{Y}) + \alpha E_f(\tilde{Y}) + \beta E_s(\tilde{Y}) \tag{1}$$

where each individual term will be explained below. α and β are parameters controlling the importance of each term.

a) Ground-truth Consistency Term E_g: The image displacements of the training locations, which are the left part of $\tilde{Y}$, should be consistent with the ground-truth. We define the penalty of violation as:

$$E_g(\tilde{Y}) = \frac{1}{N}\|Y - Y^{\mathrm{GT}}\|_F^2 = \frac{1}{N}\|\tilde{Y}P - Y^{\mathrm{GT}}\|_F^2 \tag{2}$$

where $P = [\begin{smallmatrix}\mathbf{I}_{N\times N}\\ \mathbf{0}_{N'\times N}\end{smallmatrix}]$ is the matrix selecting the left part of $\tilde{Y}$.

b) Feature Proximity Consistency Term E_f: The ith column of $\tilde{Y}$, $\mathrm{col}_i(\tilde{Y})$, encodes the output of the ith location (either a training or a test location). We construct a binary similarity matrix $S \in \{0,1\}^{(N+N')\times(N+N')}$, where $S_{ij} = 1$ if and only if the ith and jth locations are close *in the feature space*[1]. A natural assumption is that locations with similar features should have similar outputs:

$$E_f(\tilde{Y}) = \frac{1}{\sum_{i\neq j} S_{ij}} \sum_{i\neq j} S_{ij}\|\mathrm{col}_i(\tilde{Y}) - \mathrm{col}_j(\tilde{Y})\|_F^2 \tag{3}$$

For each pair of locations (i,j), E_f introduces a high penalty if they are similar in the feature space (i.e., $S_{ij} = 1$) but the output are very different (i.e., $\|\mathrm{col}_i(\tilde{Y}) - \mathrm{col}_j(\tilde{Y})\|$ is big).

The edges in S may connect two training locations, two test locations, or a training location and a test location. In this way, the ground-truth output can be propagated to the test data based on feature proximity. Denoting L_S as the Laplacian matrix of S, we can write:

$$E_f(\tilde{Y}) = \mathrm{Tr}\left(\tilde{Y}L_S\tilde{Y}^\top\right) \tag{4}$$

c) Pairwise Geometric Constraint E_s: y'_i and y'_j are displacements from two test points c'_i and c'_j to the (unknown) disc center location. From triangle geometry we have $y'_i - y'_j = c'_j - c'_i$. Therefore, we want to minimize:

$$E_s^{i,j}(Y') = \|(y'_i - y'_j) - (c'_j - c'_i)\|_2^2 = \|Y'u_{i,j} - \Delta c_{j,i}\|_F^2 \tag{5}$$

where $u_{i,j}$ is a N' dimensional vector whose ith element is 1, jth element is -1, and all others are 0s, and $\Delta c_{j,i} = c'_j - c'_i$. Adding these constraints together:

$$\begin{aligned} E_s(\tilde{Y}) &= \frac{1}{N'(N'-1)}\sum_{i\neq j} E_s^{i,j}(Y') \\ &= \frac{1}{N'(N'-1)}\|\tilde{Y}QU - \Delta C\|_F^2 \end{aligned} \tag{6}$$

where $U = [\dots, u_{i,j}, \dots]$ and $\Delta C = [\dots, \Delta c_{j,i}, \dots]$ are matrices of column vectors.

Optimization: Substituting (2), (4) and (6) into (1), we get our complete objective function as:

$$\begin{aligned} E(\tilde{Y}) = &\frac{1}{N}\,\mathrm{Tr}\left((\tilde{Y}P - Y^{\mathrm{GT}})(\tilde{Y}P - Y^{\mathrm{GT}})^\top\right) \\ &+ \alpha\,\mathrm{Tr}\left(\tilde{Y}L_S\tilde{Y}^\top\right) \\ &+ \frac{\beta}{N'(N'-1)}\,\mathrm{Tr}\left(\left(\tilde{Y}QU - \Delta C\right)\left(\tilde{Y}QU - \Delta C\right)^\top\right) \end{aligned} \tag{7}$$

We can prove that (7) is convex, and its gradient is given by (please see the supplementary material for details):

$$\begin{aligned} \frac{\partial E(\tilde{Y})}{\partial \tilde{Y}} = &\,2\tilde{Y}\left(\frac{1}{N}PP^\top + \alpha L_S + \beta\frac{1}{N'(N'-1)}QUU^\top Q^\top\right) \\ &- \frac{2}{N}Y^{\mathrm{GT}}P^\top - \frac{2\beta}{N'(N'-1)}\Delta CU^\top Q^\top \end{aligned} \tag{8}$$

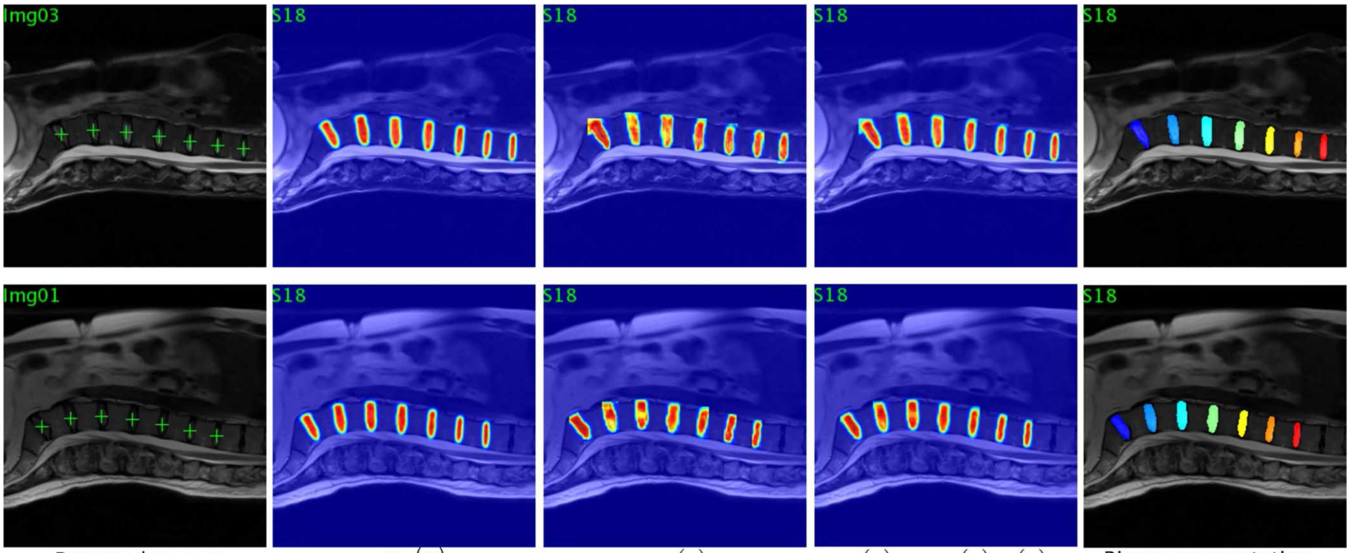
To solve for the globally optimal $\tilde{Y}$, we can either solve the equation $(\partial E(\tilde{Y}))/(\partial\tilde{Y}) = 0$ in closed form, or use gradient descent.

Discussion: E_g ensures the consistence with the ground-truth data. E_f propagates outputs from training data to test data based on feature proximity. In E_s we exploit pairwise geometric constraints to regularize the displacement values being estimated, which are not exploited in other methods such as Random Forest regression. These MRF-like neighboring constraints are encoded compactly in our objective function which can be solved globally.

Constructing Probability Maps: After we find the optimum $\tilde{Y}$, we have $Y' = \tilde{Y}Q$, and $\{c'_i + y'_i\}_{i=1\dots N'}$ will be the set of votes for the disc center position. We write $\mu_i = c'_i + y'_i$ as the position vote made by the ith patch. For each vote, there is also an uncertainty Σ_i, which is calculated as the (diagonal) variance of the training displacements that are linked to the ith test patch when we calculated the feature proximity S above. Now we calculate the probability of disc center on different image locations. We view each vote (μ_i, Σ_i) as a Gaussian distribution

---

[1] The distance between two image locations in the feature space is calculated by the Euclidean distance between the visual feature vectors at these two locations. We use knn to construct this feature similarity matrix. $S_{ij} = 1$ if an only if the ith feature is one of the k nearest neighbors of the jth feature, or vice versa.



Detected centers (a) $p_p(z)$ (b) $p_o(z)$ (c) $p(z) = p_p(z)p_o(z)$ (d) Binary segmentation (e)

Fig. 3. Overview of our IVD segmentation method. The process of two images are shown in two rows. (a): the detected disc centers; (b): the spatial probability $p_p(z)$; (c): the observation probability $p_o(z)$; (d): the combined probability $p(z)$; (e): the binary segmentation result, where the binary masks of each IVD are superimposed in the image and color is used to differentiate the IVD identification.

$G(\cdot|\mu_i, \Sigma_i)$. Then, the probability of disc center at an image coordinate (u, v, w) is given by accumulating the contribution of all votes on this image location:

$$I(u, v, w) = \sum_{i=1}^{N'} G((u, v, w)|\mu_i, \Sigma_i) \tag{9}$$

$I(u, v, w)$ is then the probability map, as in Fig. 2(d).

C. Post-Processing by Dynamic Programming

Assuming that we are interested in L IVDs, so far, we have computed L probability maps, denoted by $I_i(u, v, w)_{i=1\ldots L}$, which are derived independently for each IVD. In this step, we regularize the result by considering inter-IVD relations.

Since the spine is in a cord structure with sequential IVDs, we exploit the relative position of adjacent IVDs. For the ith and the $(i+1)$th IVDs, we collect the relative offset of their centers from the training images, and approximate the offsets by a Gaussian distribution $G_{i,i+1}(\cdot|\mu_{i,i+1}, \Sigma_{i,i+1})$ with mean $\mu_{i,i+1}$ and variance $\Sigma_{i,i+1}$. Then, the transitional probability of two IVD center positions on the test image is given by:

$$p_{i,i+1}(z_{i+1}|z_i) = G_{i,i+1}(z_{i+1} - z_i|\mu_{i,i+1}, \Sigma_{i,i+1}) \tag{10}$$

where z_i and z_{i+1} are the center positions of the ith and $i+1$th discs, respectively.

On the other hand, the observation probability is simply given by the probability map:

$$p_i(z_i) = I_i(z_i) \tag{11}$$

The optimal sequence of disc center positions are thus given by maximizing the following joint probability:

$$\text{argmax}_{z_1\ldots z_L} p_1(z_1)p_{1,2}(z_2|z_1)\ldots p_{L-1,L}(z_L|z_{L-1})p_L(z_L) \tag{12}$$

This can be solved by dynamic programming on the image grids.

III. IVD SEGMENTATION BY DATA-DRIVEN PIXEL CLASSIFICATION

A. Overview

The segmentation of a disc is performed after its center is detected. The process is illustrated in Fig. 3, where we show the segmentation process of $L = 7$ IVDs. To save space, in Fig. 3 we superimpose the visualization of the 7 discs on a single image, but please note that the segmentation is conducted separately for each disc. For each disc, suppose that the detected center location is $z_0 = (u_0, v_0, w_0)$, our task is to classify the image locations around the disc center as foreground (disc) or background. For each image location $z = (u, v, w)$, we compute two probabilities of it being the foreground: $p_p(z)$, the spatial probability, and $p_o(z)$, the observation probability. $p_p(z)$ is the probability of a pixel being the foreground given its spatial offset from the disc center $z - z_0$, which is estimated using the parzen window method from annotated training data. On the other hand, $p_o(z)$ is calculated by the data-driven method described in the next subsection based on the pixel appearance. The final probability of each pixel being a disc pixel is then given by $p(z) = p_p(z)p_o(z)$. The final binary segmentation is derived by thresholding the probability map, filling holes, and only keeping the largest connected component.

Note that $p_p(z)$ is much cheaper to calculate and serve as a good pre-filter of the potential foreground pixels. Therefore, we first calculate $p_p(z)$ over all pixels around the disc center, and then we only consider pixels where $p_p(z)$ is not zero, on which $p_o(z)$ is then calculated by the data-driven approach.

Below we explain how to calculate the observation probability p_o.

B. Calculating the Observation Probability p_o

First, we assume that training data are available. We randomly sample some image locations on those training images where the discs are already segmented. For each such image location, its ground-truth probability value $y^{\mathrm{GT}} \in \{0,1\}$ is set to 1 if it is a foreground pixel and 0 otherwise. We also calculate the feature vector of this image location, which is denoted as x. In this way, we collect a set of N training data $\{x_i, y_i\}_{i=1\ldots N}$ which consists of some feature vectors and corresponding foreground probabilities. On the test image, for each image location $z = (u,v,w)$, we calculate its feature vector x', and now we need to estimate its foreground probability, based on the training data. We note that this is a similar problem compared to the estimation of image displacement in Section II.B. The difference is that, instead of estimating the image displacement, now our estimation target is the foreground probability. We adopt a similar approach as in Section II.B, where the similar objective function is optimized with a modification in the last term, as described below.

The objective function consists of three terms:

$$E(\tilde{Y}) = E_g(\tilde{Y}) + \alpha E_f(\tilde{Y}) + \beta E_n(\tilde{Y}) \qquad (13)$$

where each individual term will be explained below. α and β are parameters controlling the importance of each term. Please note that the meaning of variables are the same as in Section II.B, except that the output value y is the foreground probability (a scalar) instead of the image displacements (in $\mathbb{R}^3$).

1. **Ground-truth Consistency Term E_g.**
   This is the same as in (2).
2. **Feature Proximity Consistency Term E_f.**
   This is the same as in (4).
3. **Neighborhood Smoothness Constraint E_n.**

The original pairwise geometric constraint E_s in (6) used in the estimation of image displacement is no longer applicable here. On the other hand, based on the nature of segmentation, we introduce a new pairwise term which enforces the neighborhood smoothness. In the case of segmentation, ${y'}_i$ is the foreground probability of the ith point. A natural assumption is that the segmentation should be smooth, i.e., neighboring points should have similar foreground probabilities. Therefore, if we define a neighboring system $\mathcal{N}$ (in this paper we use the 6-neighbor system in the 3D coordinate space), we would want to minimize:

$$E_n(\tilde{Y}) = \frac{1}{|\mathcal{N}|} \sum_{(i,j)\in\mathcal{N}} \|{y'}_i - {y'}_j\|_F^2 \qquad (14)$$

If we define A as the neighbor affinity matrix, where $A_{i,j} = 1$ iff only $(i,j) \in \mathcal{N}$, and we denote L_A as the Laplacian matrix of A, we can write E_n as:

$$E_n(\tilde{Y}) = \mathrm{Tr}\left(Y' L_A (Y')^\top\right) = \mathrm{Tr}\left(\tilde{Y} Q L_A Q^\top \tilde{Y}^\top\right) \qquad (15)$$

Optimization: Substituting (2), (4) and (15) into (13), we get our complete objective function which is in a similar form as (7). Similarly to (7), we can to prove that (13) is convex, and its gradient is given by:

$$\frac{\partial E(\tilde{Y})}{\partial \tilde{Y}} = 2\tilde{Y}\left(\frac{1}{N}PP^\top + \alpha L_S + \beta Q L_A Q^\top\right) - \frac{2}{N} Y^{\mathrm{GT}} P^\top \qquad (16)$$

To solve for the globally optimal $\tilde{Y}$, we can either solve the equation $(\partial E(\tilde{Y}))/(\partial \tilde{Y}) = 0$ in closed form, or use gradient descent.

Discussion: As in Section II.B, E_g ensures the consistence with the ground-truth data, and E_f propagates outputs from training data to test data based on feature proximity. The new term E_n enforces the neighborhood smoothness to regularize the foreground probability values being estimated. This is an MRF-like neighborhood constraint and we encode it compactly in our objective function which can be solved globally.

IV. EXPERIMENTS

A. Data

We validate our method on two datasets of T2-weighted turbo spin echo MR images. The first dataset, D1, contains images from 23 patients and the second dataset, D2, contains images of another 12 patients. Each patient was scanned with 1.5 Tesla MRI scanner (Siemens Magnetom Sonata, Siemens Healthcare, Erlangen, Germany). The resolutions of these images are varying from 2 mm × 1.25 mm × 1.25 mm to 3.3 mm × 1.1875 mm × 1.1875 mm. Note that the two datasets are not from the same series and therefore cross-dataset validation can give a good estimation of the generality of our method. As a machine-learning based technique, our method has to deal with possible difference in image resolutions between images used for training and any future test image. Usually, there are two ways to cope with it. We can either resample the images to have the same resolution, or we can set our algorithm parameters (e.g., dimension of local patches, searching ranges) using physical units, which means that the parameters will change in the pixel dimension according to the image resolution. We tried both strategy in our implementation, and did not see noticeable difference in accuracy. Finally, we choose the first way and resample all the images into the same pixel spacing of 2 mm × 1.25 mm × 1.25 mm as the unified resolution, which makes the implementation simpler with unified parameter in the pixel coordinate space.

We further assume that the images are acquired in a unified patient orientation (i.e., the patients are in the same orientation with respect to the image coordinate system), although our method tolerate moderate rotation variance up to about 15 degrees. On the other hand, we do not make any assumption about the disc position, as our localization algorithm will search for the IVDs all over the image.

We checked all images and found that the 7 IVDs T11-S1 were visible in all images. Thus, in this paper we only consider 7 IVDs T11-S1. We manually annotated these discs on all images, resulting in $7 \times 35 = 245$ discs in total. The ground-truth disc centers are defined as centroids of foreground pixels. Please note that our method does not assume that the input MR images contain exactly 7 discs. In fact, since our localization is done separately for each of the 7 IVDs T11-S1, it will automatically ignore other discs if the image contains them, as no response will be generated for them.

TABLE I
AVERAGE LOCALIZATION ERROR (IN MM) WHEN D1 IS INVOLVED FOR TEST. IMAGE #19 IN EXPERIMENT 1) FAILS ON FULLY-AUTOMATIC METHOD (ERROR: 30.6 MM). WITH MANUAL INTERVENTION OF ONE DISC, THE MISTAKE IS CORRECTED (ERROR: 1.2 MM). NUMBERS IN PARENTHESES REPRESENT THE RESULT WITH THIS MANUAL INTERVENTION. FOR ALL OTHER IMAGES, RESULTS ARE ACHIEVED FULLY-AUTOMATICALLY

1) Results of D1 leave-one-out experiment													
#1	#2	#3	#4	#5	#6	#7	#8	#9	#10	#11	#12	Mean±STD	Median
2.5	2.3	1.6	3.1	1.7	1.1	1.0	1.3	1.3	0.8	1.2	1.2		
#13	#14	#15	#16	#17	#18	#19	#20	#21	#22	#23		2.8±6.5 (1.6±0.9)	1.4 (1.4)
1.9	1.2	1.8	1.6	1.4	1.0	30.6 (1.2)	1.4	2.3	1.9	1.1			
2) Results of D2 train / D1 test experiment													
#1	#2	#3	#4	#5	#6	#7	#8	#9	#10	#11	#12	Mean±STD	Median
2.5	1.8	2.6	2.5	1.8	2.0	1.5	2.2	1.3	1.8	1.4	1.9		
#13	#14	#15	#16	#17	#18	#19	#20	#21	#22	#23		1.9±1.1	1.7
1.2	1.6	1.8	2.4	1.4	1.3	1.7	2.9	3.3	1.8	1.8			

TABLE II
AVERAGE LOCALIZATION ERROR (IN MM) WHEN D2 IS INVOLVED FOR TEST. ALL RESULTS ARE ACHIEVED FULLY-AUTOMATICALLY

3) Results of D2 leave-one-out experiment													
#1	#2	#3	#4	#5	#6	#7	#8	#9	#10	#11	#12	Mean±STD	Median
2.0	2.1	1.2	2.1	1.9	1.7	1.2	1.4	2.4	1.3	1.9	2.7	1.8±1.1	1.6
4) Results of D1 train / D2 test experiment													
#1	#2	#3	#4	#5	#6	#7	#8	#9	#10	#11	#12	Mean±STD	Median
1.7	1.5	1.6	1.6	2.7	2.1	1.6	2.0	2.4	2.3	2.2	2.1	2.0±0.9	1.9

B. Implementation Details

We use the neighborhood intensity vector as the visual feature of 3D image patches, both in disc localization and segmentation. Specifically, we draw a cube (of edge size 3 cm for localization and 1 cm for segmentation) centered on the point. The cube is then evenly divided into $4 \times 4 \times 4$ blocks, and the mean intensities in each block are concatenated to form a 64 dimensional feature and normalized to unit L_1 norm. In designing the visual feature, we are aware of the work on inter-scan MR image intensity scale standardization [28] as well as intra-scan intensity inhomogeneity correction or bias correction [29], [30] and their applications in the computer-aided assessment of anomalies in the scoliotic spine [31]. However, considering the relatively small imaging field of view in our study and the fact that the bias field is said to be smooth and slowly varying and is composed of low frequency components only [31], we choose to normalize our feature to accommodate for both intra-scan and inter-scan intensity variations: our feature vector is the concatenation of mean image intensities in different blocks within a local neighborhood (3D image volume), and then we divide the vector by its L_1 norm to make it sum up to one. This makes the feature insensitive to global or low frequency local intensity shifting, because the feature vector is not dependent on the absolute intensity in the neighborhood and what matters is the relative difference of intensities in different blocks. This makes our feature sensitive to gradient rather than the absolute intensity value.

In both localization and segmentation, for each IVD we sample about $N = 300$ patches per training image. For localization, the number of test patches $N' = 2000$. We fix $\alpha = 0.01, \beta = 0.001$ in (1) and (13).

With our Matlab implementation, it takes on average around 3 minutes to complete the localization and segmentation of a test MR image (7 discs).

Please note that both localization and segmentation are done in 3D space. However, to ease visualization, the figures in the section below are presented in 2D sagittal slices.

C. Experimental Result on IVD Localization

Evaluation Experiments: We evaluated our IVD localization method in 4 experiments:

1) **D1 leave-one-out**. The 23 images in dataset D1 are evaluated using leave-one-out.
2) **D2 train/D1 test**. The 12 images in D2 are used for training, and the 23 images in D1 are used for test.
3) **D2 leave-one-out**. The 12 images of D2 are evaluated using leave-one-out.
4) **D1 train/D2 test**. The 23 images in D1 are used for training, and the 12 images in D2 are used for test.

Evaluation Result: The evaluation metric is the localization error, which is the Euclidean distance between the ground-truth disc centers and the algorithm output, converted into physical unit of millimeter. Table I gives the results of experiments 1 and 2, where the D1 is involved for test purpose, and Table II gives the results of experiments 3 and 4, where D2 is involved for test. In each table, we show the disc localization error (average of the 7 discs) of each test image as well as the global mean, standard deviation and median value. We regard a case as failed when the average disc localization error is greater than 5 mm, which usually leads to a failure in the later segmentation.

From Tables I and II we can see that our fully-automatic localization method achieves good results in all cases except image #19 in experiment 1). In addition, we do not see notably deterioration in the cross-dataset validations with regard to the within-dataset leave-one-out cases, indicating that our method has a good generalization ability.

Fig. 4 shows illustration of our localization results on four test images in experiment 2) (D2 train/D1 test). For each image, we show: (a) The localization result from the method described in Section II.B, before the post-process step in Section II.C. This is the mode position taken from the probability map of each disc. (b) The localization result after the post-process step in Section II.C. (c) The probability maps of the 7 discs. In (a) and (b), the ground-truth disc locations are drawn by red crosses, and

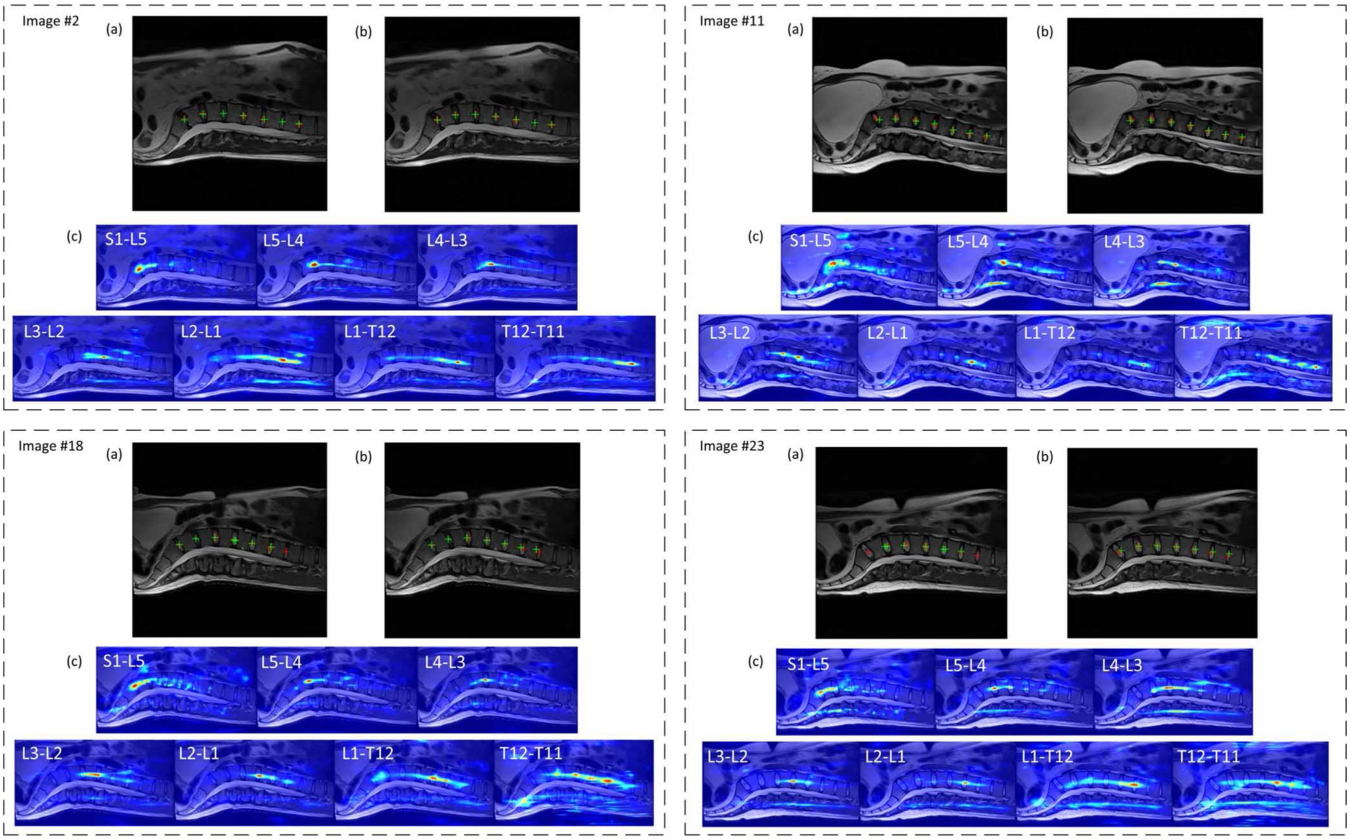


Fig. 4. Illustration of localization results of experiment 2 (D2 train/D1 test) on four test images. For each image, (a) is the localization before the final dynamic programming post-process, (b) is the result after post-process, (c) is the probability map for each of the 7 discs. (To save space, the probability maps are cropped to contain only the region near the central vertebral line).

the algorithm output locations are drawn in green. If the localization is perfect we will only see the green cross as it occludes the red one. From Fig. 4 we can see that for images #2 and #11, our data-driven estimation method already gives good result for every disc (the mode of probability maps correspond well to the ground-truth disc center locations) before the post-process. For images #18 and #23, simply taking the mode of each probability map will give some wrong positions, because the probability maps have some ambiguity due to the repetitive pattern of the discs. However, after the post-processing step which considers the global structure between the discs, we are always able to resolve this ambiguity and get correct result for every disc. We pay special attention to image #23, where the probability maps of two discs. i.e., S1-L5 and T12-T11 (the first and the last discs) are highly ambiguous and their modes coincide with their neighboring discs (e.g., the mode of the disc S1-L5 is located at the disc L5-L4 and the mode of the disc T12-T11 is located at the disc L2-L1). After the post-processing we are able to get correct localization of all discs for this image.

Failure Case and Manual Intervention: We do have one failure case as shown in Table I. When we do the leave-one-out study on D1, we are unable to get correct result for image #19 fully-automatically. The left part of Fig. 5 depicts this scenario. Our method seems to believe that the real disc S2-S1 is the disc S1-L5, the real disc S1-L5 is the disc L5-L4, …, which means that all disc centers are approximately located but the disc identifications are shifted one position to the left. If we look closely at this image, we will see that the mistake of our method is not without reason: due to the presence of lumbosacral transitional vertebrae [32], the disc space between S2 and S1 is larger than the rudimentary disc that is often seen in spines without transitions, which makes our method believe that the lumbar discs start from the real disc S2-S1. To correct this, we incorporate a manual intervention as shown in the right part of Fig. 5. In this case, we manually specify the ground-truth location of a randomly chosen disc (here the disc L1-T12) with one mouse click. We then modify the probability map of this disc to be a sharp Gaussian distribution around the ground-truth location (cf. the probability map for the disc L1-T12 in sub-figure (c)) and then run our post-processing step as before. This simple intervention on one randomly chosen disc corrects all mistakes (The mean localization error for this image is reduced from 30.6 mm to 1.2 mm).

This example of manual intervention also shows another advantage of our method. In case that the fully-automatic method fails, we only need to manually specify very few discs to correct all mistakes, and the manual intervention can be very easily integrated into our method: we only need to modify the probability map of the manually specified disc(s) as sharp Gaussians, and then run our post-processing algorithm without any change.

Compare to Other Methods: To see how our method compares to others, we also evaluate the localization using the Random Forest regression method [7]. The basic framework is similar, except that to estimate the image displacements,

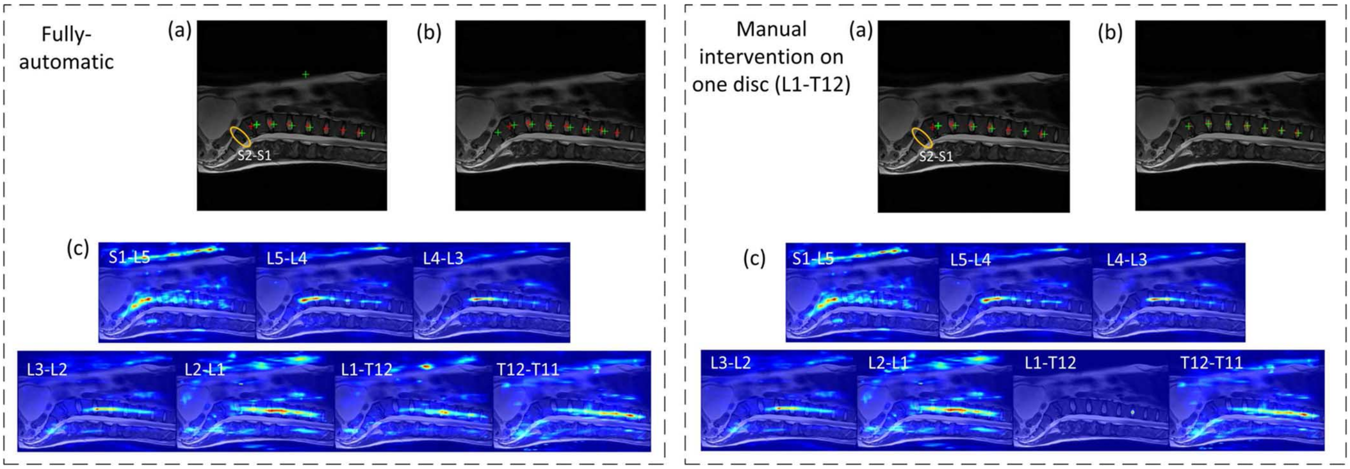


Fig. 5. Left: our fully-automatic method fails on the 19th image in the D1 leave-one-out experiment. Right: by manually annotating disc L1-T12, we are able to correctly localize all the other discs.

TABLE III
AVERAGE LOCALIZATION ERROR (IN MM) OVER ALL TEST IMAGES OF OUR APPROACH COMPARED TO RANDOM FOREST METHOD

	Our method				Random Forest regression			
Evaluation experiment	Mean±STD	Median	Min	Max	Mean±STD	Median	Min	Max
1) D1 leave-one-out	2.8±6.5	1.4	0.1	38.8	1.7±0.9	1.5	0.3	5.8
2) D2 train / D1 test	1.9±1.1	1.7	0.4	5.6	4.4±8.8	1.9	0.4	39.6
3) D2 leave-one-out	1.8±1.1	1.6	0.3	8.1	4.5±8.3	2.1	0.3	37.4
4) D1 train / D2 test	2.0±0.9	1.9	0.3	5.3	4.8±9.1	2.1	0.5	37.3

Random Forest regressor is used instead of our data-driven estimation method. The same post-process as in Section II.C is applied to the both methods. The comparison result is shown in Table III. Note that in this comparison everything is fully-automatic for both methods (i.e., we do not introduce manual intervention in case of failures). We see that our method clearly outperforms the Random Forest regression method in disc localization in all experiments. Since we cannot safely assume that the errors are distributed in a Gaussian distribution, we use Wilcoxon signed-rank test on all value pairs in Table III, which gives a $p-\text{value} = 1.9 \times 10^{-6} < 1 \times 10^{-5}$ in favor of our method. Moreover, as mentioned above, our method has only one fail case (image #19 of experiment 1)). On the other hand, Random Forest based method fails on five cases (images #3, #18, and #20 for experiment 2) and image #1 for both experiments 3) and 4)). This also shows that our method is more robust.

As a comment, we note that Kelm *et al.* [17] reported a disc localization error on MR images around 2.4 ± 1.3 mm with median value around 2.2 mm. Please note that they used a different dataset so their result is not directly comparable to ours.

D. Experimental Result on IVD Segmentation

Evaluation Experiments: We evaluated our IVD segmentation method in 2 experiments:

1) **D2 train/D1 test**. The 12 images in D2 are used for training, and the 23 images in D1 are used for test.
2) **D1 train/D2 test**. The 23 images in D1 are used for training, and the 12 images in D2 are used for test.

Note that the disc segmentation is conducted based on the disc localization output from the corresponding experiment in the previous section. For example, the output of experiment "D2 train/D1 test" in localization is used as input for the same evaluation experiment "D2 train/D1 test" in segmentation. Our segmentation algorithm is fully-automatic.

Evaluation Results: Fig. 6 shows the qualitative segmentation result of our method on 4 randomly selected cases, where (a) and (b) are from experiment 1 and (c) and (d) are from experiment 2. For each image, we show the result on three 2D sagittal slices $s-5$, s, and $s+5$, where s is the middle slice. In each slice, the red curve depicts the ground-truth disc contour and the green ones are our algorithm output. We see that our method generates visually satisfactory results.

For quantitative evaluation, we employ two metrics:

- Dice: this metric measures the percentage of correctly identified pixels, which is calculated as:

$$\text{Dice} = \frac{2|A \cap B|}{|A| + |B|} \times 100\% \tag{17}$$

  where A and B are the sets of foreground pixels in the ground-truth data and the segmentation result, respectively. Larger Dice metric means better accuracy.
- SurfDist: the average absolute distance from the ground-truth disc surface and the segmented surface. Smaller SurfDist means better accuracy.

Our method segments 3D IVDs, but as most existing methods work only on 2D sagittal slices, for comparison we calculate both the 3D and 2D versions of the two metrics explained above. The 2D measurements are computed using only the middle sagittal slice. The quantitative result of our method is shown in the first two rows in Table IV. In summary, for 3D metrics, we achieve a mean Dice of 85%–88% and a mean

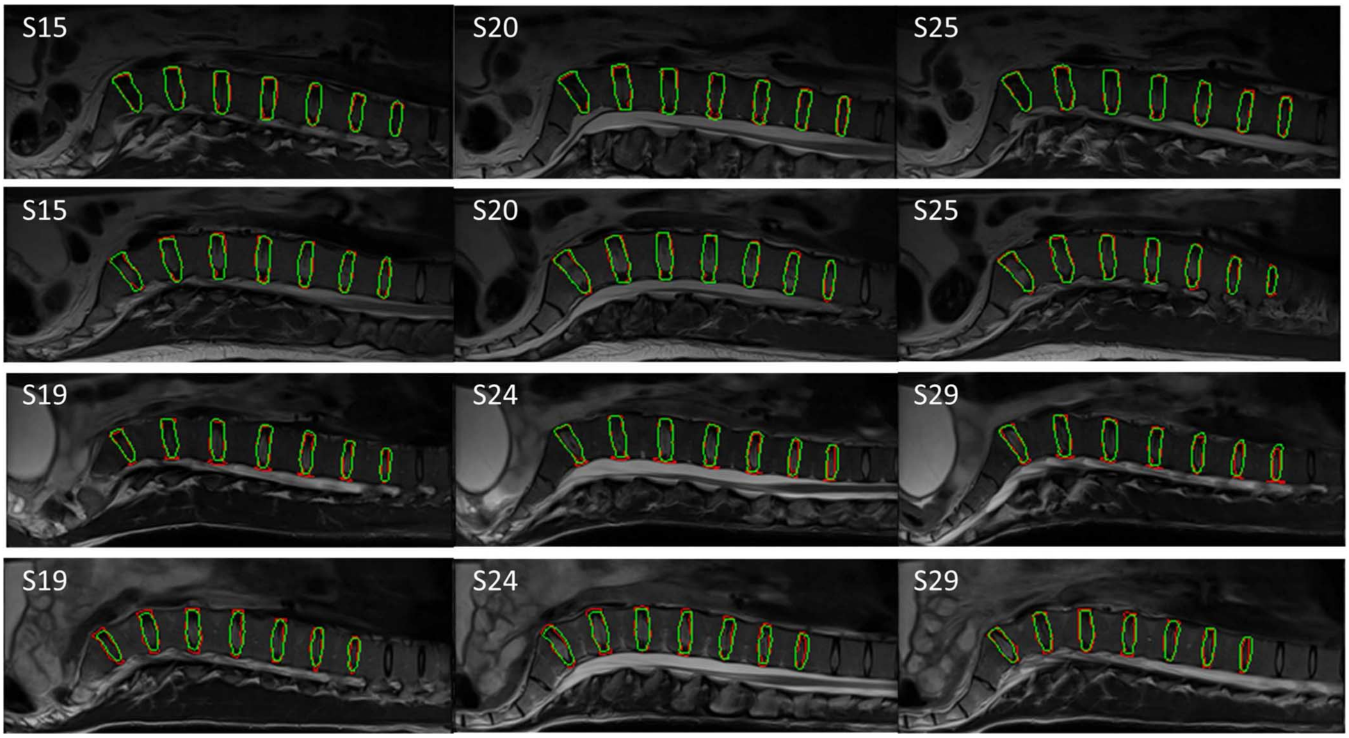


Fig. 6. Segmentation result of our method. (a): Train on D2, test on image #6 of D1; (b): Train on D2, test on image #23 of D1; (c): Train on D1, test on image #3 of D2; (d): Train on D1, test on image #10 of D2. To save space, the images are cropped to contain only the region near the central vertebral line.

TABLE IV
QUANTITATIVE EVALUATION OF OUR SEGMENTATION ALGORITHM(OURS) AND THE RF BASED METHOD

Method	Evaluation experiment	3D				2D			
		Dice(%)		SurfDist (mm)		Dice(%)		SurfDist (mm)	
		Mean±STD	Median	Mean±STD	Median	Mean±STD	Median	Mean±STD	Median
Ours	1) D2 train / D1 test	88%±4%	88%	1.3±0.2	1.3	90%±4%	91%	0.8±0.3	0.7
Ours	2) D1 train / D2 test	85%±4%	86%	1.4±0.1	1.4	87%±5%	88%	1.0±0.3	1.0
RF	1) D2 train / D1 test	85%±5%	85%	1.5±0.4	1.5	87%±5%	87%	1.1±0.5	1.0
RF	2) D1 train / D2 test	84%±4%	85%	1.5±0.2	1.4	86%±5%	87%	1.1±0.4	1.0

SurfDist of 1.3 mm–1.4 mm. For 2D metrics, we achieve a mean Dice of 87%–90% and a mean SurfDist around 0.7 mm–1.0 mm.

Compare to Other Methods: To see how our method compares to others, we also evaluate the segmentation result using the Random Forest based method. Again, the disc segmentation is conducted based on the disc localization output from the corresponding experiment in the previous section. For those failure localization cases when the Random Forest based method is used, the manual intervention introduced in the previous section is employed because otherwise it makes no sense to segment discs if disc identification is wrong. The basic segmentation framework is similar, except that to estimate the foreground probability of each pixel, Random Forest is used instead of our data-driven estimation method. The result of Random Forest based method is shown in the last two rows in Table IV. We see that our method outperforms the Random Forest based method or generates comparable results. We conducted two Wilcoxon signed-rank tests on the 3D results in Table IV, one on the 3D Dice results and the other on the 3D SurfDist results. In both cases, the tests give $p - \text{value} \ll 1 \times 10^{-5}$ in favor of our method.

As a comment, we note that Neubert *et al.* [15] reported a mean Dice of 76%–80% in their 3D IVD segmentation paper, and that in [14] Ben *et al.* reported a mean Dice of 88% in the case of 2D IVD segmentation. Please note that their experiments were conducted on different datasets so these results are not directy comparable to ours.

V. CONCLUSION

We have proposed a data-driven method for fully-automatic 3D intervertebral disc localization and segmentation of MR spine images. To localize the disc center location, we estimate the image displacements from a set of randomly sampled 3D image patches to the disc center. The displacement estimation is conducted by jointly optimizing the training and test displacement values in a data-driven way, where we take into consideration both the training data and the geometric constraint on the test image. After the disc centers are localized, we segment the discs by classifying image pixels around disc centers as background or foreground. The classification is done in a similar data-driven approach but in this case the estimation target is the foreground probability of each pixel. In addition, we introduce an extra neighborhood smooth constraint which is natural for segmentation problems. Combining the localization and segmentation method together, we provided a fully-automatic pipeline to get the IVD segmentation from MR images.

Our method is validated on T2-weighted MR images of 35 patients. We performed both within-dataset leave-one-out study and cross-dataset evaluations. Experiments show that compared to state-of-the-art, our method achieves better or comparable results. Specifically, we achieved for localization a mean error of 1.6–2.0 mm, and for segmentation a mean Dice metric of 85%–88% and a mean surface distance of 1.3–1.4 mm. During our experiment, our fully-automatic method fails on one image out of 35, and this can be easily corrected by manual clicking only one disc center location.

A limitation of our method is that we assume that the test image contains at least 7 IVDs (T11-S1). If there are more than these 7 discs in the image our algorithm still works, because other discs will not generate strong response for any of the 7 discs that we are looking for. This can be seen in several results presented in this paper, for example in Fig. 4 the image #18 contains 9 IVDs (T9-S1) and our method is not affected. However, if the image does not contain all the 7 discs (T11-S1), then our algorithm may fail, because the location of the missing discs returned by our algorithm may be completely wrong. One possible solution is to enable the algorithm to automatically judge if one or more discs are missing based on the magnitude of the corresponding detected response.

In the future, we would like to study the diagnosis of intervertebral disc degeneration based on the segmentation results of our algorithm. We also plan to integrate our localization/segmentation method on more anatomical structures. Furthermore, to promote the development of automated 3D MR image analysis methods, and to provide a standardized ground for benchmarking automated methods, we will soon release all 35 3D T2-weighted MR images used in this paper to the community via the Computational Methods and Clinical Applications for Spine Imaging workshop official website (http://spineweb.digitalimaginggroup.ca/spineweb).

REFERENCES

[1] M. T. Modic and J. S. Ross, "Lumbar degenerative disk disease," *Radiology*, vol. 245, pp. 43–61, 2007.

[2] H. An *et al.*, "Introduction. Disc degeneration: Summary," *Spine*, vol. 29, pp. 2677–2678, 2004.

[3] P. Violas, E. Estivalezes, J. Briot, J. Sales de Gauzy, and P. Swider, "Objective quantification of intervertebral disc volume properties using MRI in idiopathic scoliosis surgery," *J. Magn. Reson. Imag.*, vol. 25, pp. 386–391, 2007.

[4] R. Niemelainen, T. Videman, S. Dhillon, and M. Battie, "Quantitative measurement of intervertebral disc signal using MRI," *Clin. Radiol.*, vol. 63, pp. 252–255, 2008.

[5] S. Schmidt *et al.*, "Spine detection and labeling using a parts-based graphical model," in *Int. Conf. Inf. Process. Med. Imag.*, 2007, vol. 4584, LNCS, pp. 122–133.

[6] J. J. Corso, R. S. Alomari, and V. Chaudhary, "Lumbar disc localization and labeling with a probabilistic model on both pixel and object features," in *Proc. 11th Int. Conf. Med. Image Comput. Comput.-Assist. Intervent.*, 2008, vol. 5241, LNCS, pp. 202–210.

[7] B. Glocker, J. Feulner, A. Criminisi, D. R. Haynor, and E. Konukoglu, "Automatic localization and identification of vertebrae in arbitrary field-of-view CT scans," in *Proc. 15th Int. Conf. Med. Image Comput. Comput.-Assist. Intervent.*, 2012, vol. 7512, LNCS, pp. 590–598.

[8] B. Glocker, D. Zikic, E. Konukoglu, D. R. Haynor, and A. Criminisi, "Vertebrae localization in pathological spine CT via dense classification from sparse annotations," in *Proc. 16th Int. Conf. Med. Image Comput. Comput.-Assist. Intervent.*, 2013, vol. 8150, LNCS, pp. 262–270.

[9] D. Major, J. Hladuvka, F. Schulze, and K. Buhler, "Automated landmarking and labeling of fully and partially scanned spinal columns in CT images," *Med. Image Anal.*, vol. 17, no. 8, pp. 1151–1163, 2013.

[10] R. Shi, D. Sun, Z. Qiu, and K. L. Weiss, "An efficient method for segmentation of MRI spine images," in *IEEE/ICME Int. Conf. Complex Med. Eng.*, May 2007, pp. 713–717.

[11] C. Chevrefils, F. Cheriet, C. E. Aubin, and G. Grimard, "Texture analysis for automatic segmentation of intervertebral disks of scoliotic spines from MR images," *IEEE Trans, Inf. Technol. Biomed.*, vol. 13, no. 4, pp. 608–620, Jul. 2009.

[12] S. K. Michopoulou *et al.*, "Atalas-based segmentation of degenerated lumbar intervertebral discs from MR images of the spine," *IEEE Trans. Biomed. Eng.*, vol. 56, no. 9, pp. 2225–2231, Sep. 2009.

[13] S.-H. Huang, Y.-H. Chu, S.-H. Lai, and C. L. Novak, "Learning-based vertebra detection and iterative normalized-cut segmentation for spinal MRI," *IEEE Trans. Med. Imag.*, vol. 28, no. 10, pp. 1595–1605, Oct. 2009.

[14] I. Ben Ayed, K. Punithakumar, G. Garvin, W. Romano, and S. Li, "Graph cuts with invariant object-interaction priors: Application to intervertebral disc segmentation," in *Int. Conf. Inf. Process. Med. Imag.*, 2011, vol. 6801, LNCS, pp. 221–232.

[15] A. Neubert *et al.*, "Automatic 3d segmentation of vertebral bodies and intervertebral discs from MRI," in *Int. Conf. Digital Imag, Comput., Tech. Appl.*, Dec. 2011, pp. 19–24.

[16] M. W. K. Law, K. Tay, A. Leung, G. J. Garvin, and S. Li, "Intervertebral disc segmentation in MR images using anisotropic oriented flux," *Med. Image Anal.*, vol. 17, pp. 43–61, 2013.

[17] B. M. Kelm *et al.*, "Spine detection in CT and MR using iterated marginal space learning," *Med. Image Anal.*, vol. 17, no. 8, pp. 1283–1292, 2013.

[18] Y. Zheng, A. Barbu, B. Georgescu, M. Scheuering, and D. Comaniciu, "Four-chamber heart modeling and automatic segmentation for 3-D cardiac CT volumes using marginal space learning and steerable features," *IEEE Trans. Med. Imag.*, vol. 27, no. 11, pp. 1668–1681, Nov. 2008.

[19] Y. Boykov and G. Funka-Lea, "Graph cuts and efficient n-d image segmentation," *Int. J. Comput. Vis.*, vol. 70, pp. 109–131, 2006.

[20] C. Chen *et al.*, "Automatic x-ray landmark detection and shape segmentation via data-driven joint estimation of image displacements," *Med. Image Anal.*, vol. 18, no. 3, pp. 487–499, 2014.

[21] A. Criminisi, J. Shotton, D. P. Robertson, and E. Konukoglu, "Regression forests for efficient anatomy detection and localization in CT studies," in *Med. Comput. Vis. 2010: Recognit. Tech. Appl. Med. Imag. MICCAI Workshop*, 2010, pp. 106–117.

[22] O. Pauly, "Fast multiple organ detection and localization in whole-body MR Dixon sequences," in *Proceedings of the 14th International Conference on Medical Image Computing and Computer-Assisted Intervention*. Berlin, Germany: Springer-Verlag, 2011, p. 239.

[23] C. Lindner, S. Thiagarajah, J. M. Wilkinson, G. A. Wallis, and T. F. Cootes, "Accurate fully automatic femur segmentation in pelvic radiographs using regression voting," in *Proc. 14th Int. Conf. Med. Image Comput. Comput.-Assist. Intervent.*, 2012, vol. 3, pp. 353–360.

[24] R. Donner, B. H. Menze, H. Bischof, and G. Langs, "Global localization of 3d anatomical structures by pre-filtered hough forests and discrete optimization," *Med. Image Anal.*, vol. 17, no. 8, pp. 1304–1314, 2013.

[25] R. Tibshirani, "Regression shrinkage and selection via the LASSO," *J. R. Stat. Soc.*, vol. 58, no. 1, pp. 267–288, 1996.

[26] T. Cao, C. Zach, S. Modla, D. Powell, and K. Czymmek, "Multi-modal registration for correlative microscopy using image analogies," *Med. Image Anal.*, vol. 18, no. 6, pp. 914–926, 2014.

[27] L. Wang *et al.*, "Automated bone segmentation from dental CBCT images using patch-based sparse representation and convex optimization," *Med. Phys.*, vol. 41, no. 4, p. 043503, 2014.

[28] F. Jäger and J. Hornegger, "Nonrigid registration of joint histograms for intensity standardization in magnetic resonance imaging," *IEEE Trans. Med. Imag.*, vol. 28, no. 1, pp. 137–150, Jan. 2009.

[29] C. Li, J. C. Gore, and C. Davatzikos, "Multiplicative intrinsic component optimization (MICO) for MRI bias field estimation and tissue segmentation," *Magn. Reson. Imag.*, vol. 32, no. 7, pp. 913–923, 2014.

[30] N. J. Tustison *et al.*, "NN4ITK: Improved N3 bias correction," *IEEE Trans. Med. Imag.*, vol. 29, no. 6, pp. 1310–1320, Jun. 2010.

[31] F. Jäger, "Normalization of magnetic resonance images and its application to the diagnosis of the scoliotic spine," Ph.D. dissertation, Univ. Erlangen-Nuremberg, Erlangen, Germany, 2011.

[32] G. P. Konin and D. M. Walz, "Lumbosacral transitional vertebrae: Classification, imaging findings, and clinical evidence," *AJNR Am J. Neuroradiol.*, vol. 31, no. 10, pp. 1778–1786, 2010.