3D Intervertebral Disc Localization and Segmentation from MR Images by Data-Driven Regression and Classification

Cheng Chen[1], D. Belavy[2], and Guoyan Zheng[1]

[1] Institute for Surgical Technology and Biomechanics,
University of Bern, Switzerland
{cheng.chen,guoyan.zheng}@istb.unibe.ch
[2] Department of Radiology, Charite University Medicine Berlin, Germany

Abstract. In this paper we propose a new fully-automatic method for localizing and segmenting 3D intervertebral discs from MR images, where the two problems are solved in a unified data-driven regression and classification framework. We estimate the output (image displacements for localization, or fg/bg labels for segmentation) of image points by exploiting both training data and geometric constraints simultaneously. The problem is formulated in a unified objective function which is then solved globally and efficiently. We validate our method on MR images of 25 patients. Taking manually labeled data as the ground truth, our method achieves a mean localization error of 1.3 mm, a mean Dice metric of 87%, and a mean surface distance of 1.3 mm. Our method can be applied to other localization and segmentation tasks.

1 Introduction

In clinical practice, accurate identifying of intervertebral discs (IVD) is very important for diagnosis and operation planning of spine pathologies. In this paper we propose a fully automatic method to localize and segment 3D IVDs from MR image with a unified regression and classification framework.

In literature, different methods have been proposed for IVD localization [1,2] and segmentation [5,6,7,8,9]. In [1], the IVDs were localized and labeled by a probabilistic model considering image intensity and geometric constraints. Corso et al. [2] enforced the inter-disc distance constraint to improve the label accuracy. Glocker et al. applied the Random Forest regression [3] and classification [4] methods, although their localization target is the vertebrae instead of IVD.

For IVD segmentation, existing methods are based on watershed algorithm [5], atlas registration [6], graph cuts with geometric priors from neighboring discs [7], template matching and statistic shape model [8], or anisotropic oriented flux detection [9]. All of these methods except [8] work only on 2D sagittal images.

Recently, a new data-driven optimization method [10] was proposed for landmark localization. Inspired by this, in this paper we make four contributions. (1): We extend the method into segmentation domain, where we estimate the

G. Wu et al. (Eds.): MLMI 2014, LNCS 8679, pp. 50–58, 2014.

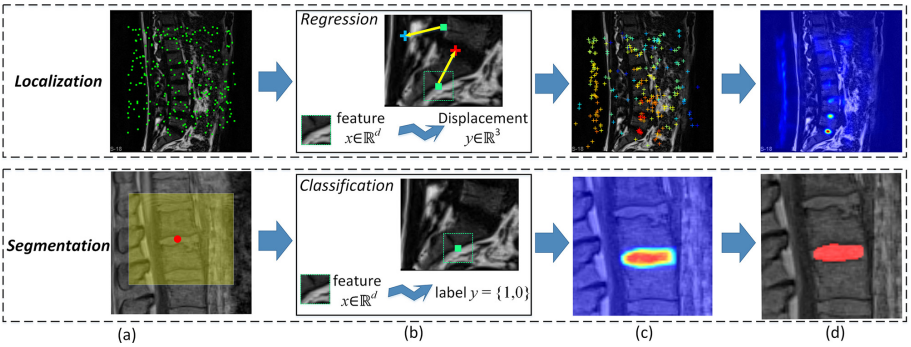

Fig. 1. Pipeline overview our method. Top: localization. Bottom: segmentation.

foreground/background label of image points instead of displacements. (2): We introduce a new constraint for segmentation which ensures the neighborhood smoothness. (3): We unify our localization and segmentation solutions into one unified framework, where we estimate output values (displacements or labels) on image locations. (4): We verified our method on MR images.

2 Data-Driven Regression/Classification Method

2.1 Overview

The localization and segmentation problems are formulated as in Fig. 1. Given an image, we consider a set of points (Fig. 1(a)): for localization task these are some randomly sampled points (green dots), and for segmentation these are all voxels inside a region of interest (yellow box). Each of these points can be represented by its visual feature calculated in a small image neighborhood (the green dash box in Fig. 1(b)). Then, we want to estimate the output values for each point. In the case of localization, the output is the displacement vector from the point to the target position (e.g. disc center), which makes it a *regression* problem. Each point makes a vote relative to itself (Fig. 1(c)) and a score map can be estimated by aggregating these votes (Fig. 1(d)). For segmentation, we estimate the fg/bg label of each voxel (Fig. 1(c)), which is a soft *classification* problem. The binary segmentation is then derived from the soft labels (Fig. 1(d)).

Notations. Suppose that N points are sampled on the training images, and let $\{x_i\}_{i=1...N}$ denote the features calcuated at these points, where $x_i \in \mathbb{R}^d$. We denote $X = [x_1...x_N] \in \mathbb{R}^{N \times d}$. We use $\{y\}_{i=1...N}$ to denote the output value of the training points, i.e. $y_i \in \mathbb{R}^3$ for localization, and $y_i \in \{0, 1\}$ for segmentation. The training images are annotated, so that the ground-truth output values of training points are known as $\{y_i^{GT}\}_{i=1...N}$, and we denote $Y^{GT} = [y_1^{GT}...y_N^{GT}]$.

Given a new image, we randomly sample N' points at locations $\{c'\}_{i=1...N'}$, whose features are $\{x'\}_{i=1...N'}$. We denote $X' = [x'_1...x'_{N'}]$. The task is to compute the output values for these points $\{y'\}_{i=1...N'}$. We write $Y' = [y'_1...y'_{N'}]$.

We solve for Y' by optimizing an objective function as below. Please refer to the supplementary material for a complete mathematical treatment.

2.2 Objective Function

First, we construct a matrix $\tilde{Y} = [Y, Y']$ which is the composition of training and test outputs. 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 $Y' = \tilde{Y}Q$, where Q is a $\begin{pmatrix}\mathbf{0} & \mathbf{1}\end{pmatrix}^T$ matrix selecting the right part.

1. Ground-truth Consistence E_g. The output of the training points, which is the left part of $\tilde{Y}$, should be consistent with the ground-truth. With a (0,1) matrix P selecting the left part of $\tilde{Y}$, we define the penalty of violation as:

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

2. Feature Proximity Consistence E_f. The ith column of $\tilde{Y}$, $\mathrm{col}_i(\tilde{Y})$, encodes the output of the ith point (either a training or a test point). We construct a binary similarity matrix $S \in \{0,1\}^{(N+N')\times(N+N')}$, where $S_{ij} = 1$ iff the ith and jth points are mutually k nearest neighbors *in the feature space*. A natural assumption is that points with similar features should have similar outputs:

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

For each pair of points (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). Denoting L_S as the Laplacian matrix of S, we can write:

$$E_f(\tilde{Y}) = \mathrm{Tr}\left(\tilde{Y}L_S\tilde{Y}^\top\right) \tag{3}$$

3. Point Subtractive Constraint E_s. In the case of localization, y'_i and y'_j are displacements from two test points c'_i and c'_j to the (unknown) target location. From triangle geometry we have $y'_i - y'_j = c'_j - c'_i$. Therefore, we want to minimize:

$$E_s^{i,j}(Y') = \|(y'_i - y'_j) - (c'_j - c'_i)\|_2^2 = \|Y'u_{i,j} - \Delta c_{j,i}\|_F^2 \tag{4}$$

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:

$$E_s(\tilde{Y}) = \frac{1}{N'(N'-1)} \sum_{i\neq j} E_s^{i,j}(Y') = \frac{1}{N'(N'-1)}\|\tilde{Y}QU - \Delta C\|_F^2 \tag{5}$$

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

4. Point Neighborhood Constraint E_n. In the case of segmentation, y'_i is the label of the ith point. A natural assumption is that the segmentation should

be smooth, i.e. neighboring points should have similar labels. Therefore, if we define a neighboring system $\mathcal{N}$, we would want to minimize:

$$E_n(\tilde{Y}) = \frac{1}{|\mathcal{N}|} \sum_{(i,j)\in\mathcal{N}} \|y'_i - y'_j\|_F^2 \tag{6}$$

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:

$$E_n(\tilde{Y}) = \mathrm{Tr}\left(Y' L_A (Y')^\top\right) = \mathrm{Tr}\left(\tilde{Y} Q L_A Q^\top \tilde{Y}^\top\right) \tag{7}$$

The Objective Function. Our objective function consists of the above terms:

$$E(\tilde{Y}) = E_g(\tilde{Y}) + \alpha E_f(\tilde{Y}) + \beta E_s(\tilde{Y}) + \gamma E_n(\tilde{Y}) \tag{8}$$

where the terms are defined in Eqs. (1), (3), (5) and (7), with their respective importance controlled by parameters α, β and γ. Note that E_s is defined only for localization ($\gamma = 0$), and E_n is only defined for segmentation ($\beta = 0$).

Optimization. Without loss of generality, we relax the binary requirement of labels in the segmentation case, and let labels y to be continuous. It is not difficult to prove that Eq. (8) is convex, with gradient given by:

$$\begin{aligned}\frac{\partial E(\tilde{Y})}{\partial \tilde{Y}} = \tilde{Y}\left(\frac{1}{N} P P^\top + \alpha L_S + \beta \frac{1}{N'(N'-1)} Q U U^\top Q^\top + \gamma Q L_A Q^\top\right) \\ - \frac{1}{N} Y^{GT} P^\top - \frac{\beta}{N'(N'-1)} \Delta C U^\top Q^\top\end{aligned} \tag{9}$$

For the globally optimal $\tilde{Y}$, we can either solve the equation $\frac{\partial E(\tilde{Y})}{\partial \tilde{Y}} = 0$ in closed form, or use gradient descent from the initialization given by k-nn search.

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. The key contribution is that in E_s and E_n we exploit different pairwise geometric constraints to regularize the output values being estimated, which are not exploited in other methods, such as [3]. These MRF-like neighboring constraints are encoded compactly in our objective function which can be solved globally.

3 Application to IVD Localization and Segmentation

We applied our method to IVD, where we first localize the disc centers, and then segment the discs. Without loss of generality, we consider 7 discs T11-L5 and number them reversely from 1 (L5) to 7 (T11). Note that for both localization and segmentation, the training and prediction are done separately for each IVD, which means that the presence of other IVDs outside T11-L5 will not affect our method as those IVDs will not generate significant response.

Localization of disc centers

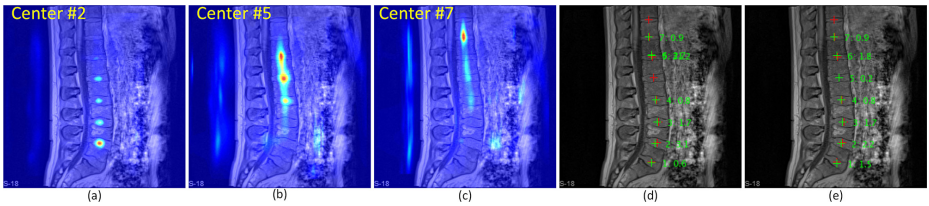

Fig. 2. The first step of localization. (a)-(c): Score images of three disc centers 2, 5 and 7. (d): The mode of each score image. (e): After HMM optimization. For (d) and (e), the red crosses are ground-truth center locations and the greens are detected centers.

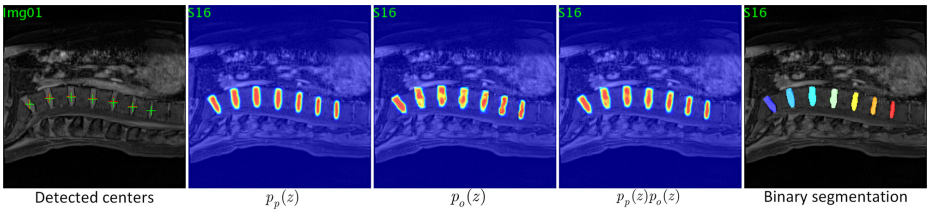

Fig. 3. The segmentation process after the disc centers are detected

For each disc center, the method in Section 2 will sample a set of points over the image and produce a set of votes. We aggregate these discrete votes to produce a continuous soft score map by considering each vote as a small Gaussian distribution [10]. Therefore, for each image, 7 score maps are produced.

We detect the disc centers in a two-step coarse-to-fine way. In the first step, points are sampled over the entire image to search for the disc centers, as in Fig. 2. Due to the repetitive pattern, the produced score maps are multimodal with potential ambiguities. For example, in Fig. 2(d) the center 5 is confused with center 6 if we simply take mode of its score map. To improve the robustness, the score maps are treated as observation probabilities and are fed to an HMM model encoding the prior geometric information of neighboring disc centers as in [3]. In the second step, we fine-tune the center locations by sampling points only in a local region around the centers initialized from the first step.

Segmentation of Discs

The segmentation of a disc is performed after its center is detected at location $z_0 = (u_0, v_0, w_0)$. The process is shown in Fig. 3. To save space, we superimpose the visualization of the 7 discs on a single image, but the segmentation is conducted separately for each disc. For each pixel location $z = (u, v, w)$, we compute two probabilities of it being the foreground of a disc: $p_p(z)$, the prior probability, and $p_o(z)$, the observation probability. $p_p(z)$ is the probability of being the foreground given the offset from the disc center $z - z_0$, which is estimated using the parzen window method from the annotated training data. On the other hand, $p_o(z)$ is calculated by the data-driven estimation method in Section 2. Since $p_p(z)$ is much cheaper to calculate and serve as a good pre-filter

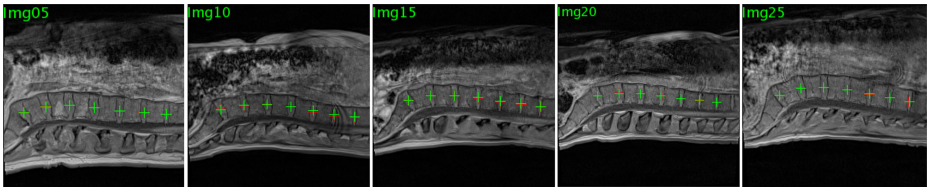

Fig. 4. The qualitative localization result on some images (the 18th sagittal slice)

Table 1. Quantitative evaluation of disc center localization

	Median	Mean	Std.	Min.	Max.
Ours	1.3	1.3	0.6	0.2	3.0
Random Forest [3]	1.6	2.7	6.2	0.3	40.6

of the potential foreground pixels, we first calculate $p_p(z)$ over all pixels around the disc center, and then we only consider voxels where $p_p(z)$ is not zero, on which $p_o(z)$ is then calculated. The final probability of each pixel is then given by $p(z) = p_p(z)p_o(z)$. The final binary segmentation is derived by thresholding the probability map and only keeping the largest connected component.

4 Experiments

Data

We validate our method on MR images of 25 patients. Each patient was scanned with 1.5 Tesla MRI scanner of Siemens. Dixon protocol was used to reconstruct four aligned high-resolution 3D volumes during one data acquisition: in-phase, opposed-phase, fat and water images. We manually annotated the intervertebral discs in water images of all subjects, resulting in 175 discs in total. The ground-truth disc centers are defined as disc centroids. The study is conducted in a leave-one-out manner. In each round data of 1 subject is chosen for testing and data of the remaining 24 subjects are used for training purpose.

Implementation Details

We use the neighborhood intensity vector as the visual feature of sampled image points. Specifically, we draw a cube (of edge size 3cm for localization and 1cm 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. As our data contains 4 channels, we concatenate the vector from all channels to form a 256 dimensional final feature vector. For parameter selection, we fix $\alpha = 0.01, \beta = 0.001, \gamma = 0$ for localization, and $\alpha = 0.01, \beta = 0, \gamma = 0.01$ for segmentation. Our unoptimized Matlab implementation requires on average 3.5 minutes to finish both localization and segmentation of one subject. Please note that all our operations are done in 3D space. However, to ease visualization, the figures in the following sections are presented in 2D sagittal slices.

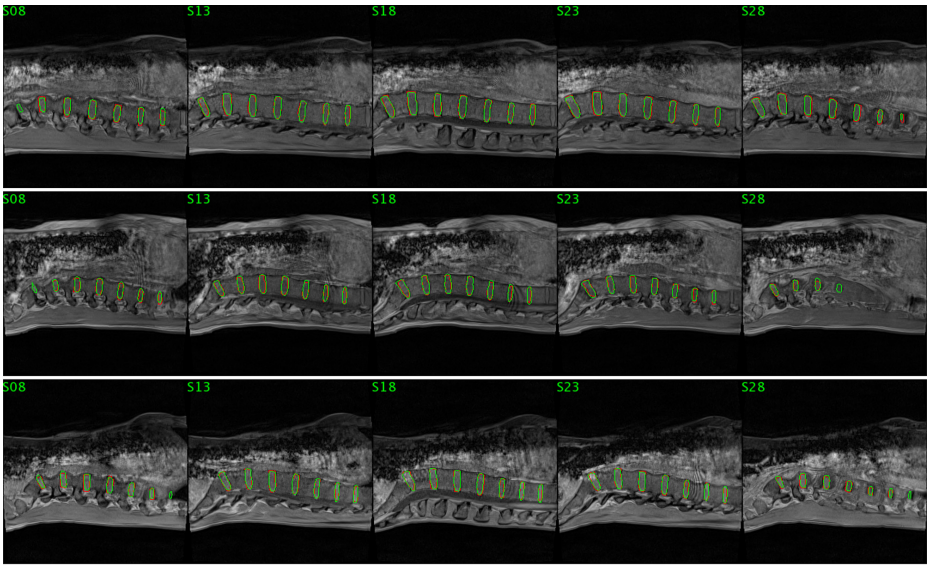

Fig. 5. Segmentation result on three images. We visualize the result on the 8th, 13th, 18th, 23th, 28th sagittal slices. Red: ground-truth contour. Green: our results.

Localization Result

Fig. 4 shows some qualitative results of disc center localization (only the 18th sagittal slice is shown), where the red crosses are ground-truth and the green ones are the detected centers. We also conducted quantitative evaluation as in Table 1, where the evaluation metric is the Euclidean distance from the detected disc centers to the ground-truth. We get a mean localization error of 1.3mm. We also compare our results with the Random Forest based method [3]. To make the comparison fair, we use the same parameters (e.g. the same features...) and the same HMM optimization process for both methods. From the result we can see that we do get better results.

Segmentation Result

We show our qualitative segmentation result on randomly selected three images in Fig. 5. We visualize the results by superimposing the contours of ground-truth discs and those of our results on five sagittal slices (slices 8,13,18,23 and 28). The red contours are ground-truth and the green ones are our results.

For quantitative evaluation, we employ two metrics: the Dice metric which measures the percentage of correctly identified pixels, and the average physical distance from the ground-truth disc surface and the segmented surface. The results are summarized in Table 2. We achieve a mean Dice of 87% and a mean SurfDist of 1.3mm. We note that Neubert et al. [8] reported a mean Dice of 76%-80% in their 3D IVD segmentation paper on a different dataset.

Table 2. Quantitative evaluation of disc segmentation. The unit of SurfDist is mm.

	Median	Mean	Std.	Min.	Max.
Dice (3D)	87%	87%	3%	76%	92%
SurfDist (3D)	1.3	1.3	0.2	1.0	2.4
Dice (sagittal)	91%	90%	4%	72%	96%
SurfDist (sagittal)	0.7	0.7	0.3	0.3	1.6

Since most existing methods work only on 2D sagittal slices, for comparison we also calculate the 2D versions of the metrics by using only the 18th slice (in most cases it is the centered sagittal slice), where we achieve a mean Dice of 90% and SurfDist of 0.7mm. We note that in [7] they reported a mean Dice of 88% in the case of 2D IVD segmentation on a different dataset.

5 Conclusions

We have proposed a unified framework for localization and segmentation tasks of medical images. We estimate outputs (displacements or labels) on image points by considering both training data and geometric constraints. Applied to the intervertebral disc case on MR data, our method achieves good results. Our method can be generally applied to other localization and segmentation tasks, and in the future, we plan to conduct more studies on different types of images.

References

1. Schmidt, S., Kappes, J.H., Bergtholdt, M., Pekar, V., Dries, S.P.M., Bystrov, D., Schnörr, C.: Spine detection and labeling using a parts-based graphical model. In: Karssemeijer, N., Lelieveldt, B. (eds.) IPMI 2007. LNCS, vol. 4584, pp. 122–133. Springer, Heidelberg (2007)
2. Corso, J.J., Alomari, R.S., Chaudhary, V.: Lumbar disc localization and labeling with a probabilistic model on both pixel and object features. In: Metaxas, D., Axel, L., Fichtinger, G., Székely, G. (eds.) MICCAI 2008, Part I. LNCS, vol. 5241, pp. 202–210. Springer, Heidelberg (2008)
3. Glocker, B., Feulner, J., Criminisi, A., Haynor, D.R., Konukoglu, E.: Automatic localization and identification of vertebrae in arbitrary field-of-view CT scans. In: Ayache, N., Delingette, H., Golland, P., Mori, K. (eds.) MICCAI 2012, Part III. LNCS, vol. 7512, pp. 590–598. Springer, Heidelberg (2012)
4. Glocker, B., Zikic, D., Konukoglu, E., Haynor, D.R., Criminisi, A.: Vertebrae localization in pathological spine CT via dense classification from sparse annotations. In: Mori, K., Sakuma, I., Sato, Y., Barillot, C., Navab, N. (eds.) MICCAI 2013, Part II. LNCS, vol. 8150, pp. 262–270. Springer, Heidelberg (2013)
5. Chevrefils, C., Cheriet, F., Aubin, C.E., Grimard, G.: Texture analysis for automatic segmentation of intervertebral disks of scoliotic spines from mr images. IEEE Trans. on Information Technology in Biomedicine 13, 608–620 (2009)

6. Michopoulou, S.K., Costaridou, L., Panagiotopoulos, E., Speller, R., Panayiotakis, G., Todd-Pokropek, A.: Atalas-based segmentation of degenerated lumbar intervertebral discs from mr images of the spine. IEEE Trans. on Biomedical Engineering 56(9), 2225–2231 (2009)
7. Ben Ayed, I., Punithakumar, K., Garvin, G., Romano, W., Li, S.: Graph cuts with invariant object-interaction priors: Application to intervertebral disc segmentation. In: Székely, G., Hahn, H.K. (eds.) IPMI 2011. LNCS, vol. 6801, pp. 221–232. Springer, Heidelberg (2011)
8. Neubert, A., Fripp, J., Shen, K., Salvado, O., Schwarz, R., Lauer, L., Engstrom, C., Crozier, S.: Automatic 3D segmentation of vertebral bodies and intervertebral discs from mri. In: International Conference on Ditial Imaging Computing: Techniques and Applications (2011)
9. Law, M.W.K., Tay, K., Leung, A., Garvin, G.J., Li, S.: Intervertebral disc segmentation in mr images using anisotropic oriented flux. Medical Image Analysis 17, 43–61 (2013)
10. Chen, C., Xie, W., Franke, J., Grutzner, P.A., Nolte, L.-P., Zheng, G.: Automatic x-ray landmark detection and shape segmentation via data-driven joint estimation of image displacements. Medical Image Analysis 18, 487–499 (2014)