Computerized Medical Imaging and Graphics 38 (2014) 639–649

Contents lists available at ScienceDirect

Computerized Medical Imaging and Graphics

journal homepage: www.elsevier.com/locate/compmedimag

Supervised methods for detection and segmentation of tissues in clinical lumbar MRI

Subarna Ghosh*, Vipin Chaudhary

Department of Computer Science and Engineering, State University of New York (SUNY) at Buffalo, NY 14260, United States

ARTICLE INFO

Article history:
Received 16 November 2013
Received in revised form 14 March 2014
Accepted 21 March 2014

Keywords:
Clinical lumbar MRI
Computer-aided diagnosis
Intervertebral disc segmentation
Dural sac segmentation

ABSTRACT

Lower back pain (LBP) is widely prevalent all over the world and more than 80% of the people suffer from LBP at some point of their lives. Moreover, a shortage of radiologists is the most pressing cause for the need of CAD (computer-aided diagnosis) systems. Automatic localization and labeling of intervertebral discs from lumbar MRI is the first step towards computer-aided diagnosis of lower back ailments. Subsequently, for diagnosis and characterization (quantification and localization) of abnormalities like disc herniation and stenosis, a completely automatic segmentation of intervertebral discs and the dural sac is extremely important. Contribution of this paper towards clinical CAD systems is two-fold. First, we propose a method to automatically detect all visible intervertebral discs in clinical sagittal MRI using heuristics and machine learning techniques. We provide a novel end-to-end framework that outputs a tight bounding box for each disc, instead of simply marking the centroid of discs, as has been the trend in the recent past. Second, we propose a method to simultaneously segment all the tissues (vertebrae, intervertebral disc, dural sac and background) in a lumbar sagittal MRI, using an auto-context approach instead of any explicit shape features or models. Past work tackles the lumbar segmentation problem on a tissue/organ basis, and which tend to perform poorly in clinical scans due to high variability in appearance. We, on the other hand, train a series of robust classifiers (random forests) using image features and sparsely sampled context features, which implicitly represent the shape and configuration of the image. Both these methods have been tested on a huge clinical dataset comprising of 212 cases and show very promising results for both disc detection (98% disc localization accuracy and 2.08 mm mean deviation) and sagittal MRI segmentation (dice similarity indices of 0.87 and 0.84 for the dural sac and the inter-vertebral disc, respectively).

1. Introduction

Statistics from the National Institutes of Health (NIH) shows that 70–85% of all people have back pain at some time in their life, and that back pain is the most frequent cause of activity limitation in people younger than 45 years old. According to the National Center for Health Statistics, more than 30 million MRI exams are conducted annually in the US and half of them are spine-related [1]. A staggering 50 billion dollars are spent annually on health care and rehabilitation for back related issues. The lower back or lumbar spine helps in structural support, movement and protection of body tissues. It consists of five bones called lumbar vertebrae, stacked one upon the other that connect the upper spine to the pelvis; six shock absorbers, called intervertebral discs, which act both as cushions and stabilizers to protect the lumbar vertebrae; and spinal cord and nerves, which travel through a central canal in the lumbar vertebrae, connecting our brain to the muscles of the legs. Back pain may be a symptom of many different causes, like trauma, degeneration of vertebrae, infection, abnormal growth (tumor), obesity, protruding or herniated disc, disease (i.e., osteoarthritis, spondylitis, compression fractures), etc. Fig. 1(a) shows a detailed illustration of various disc problems affecting the lumbar spine.

In the past decade, there has been a severe shortage of radiologists [2] and projections show that by the year 2020 the demand for radiologists will far exceed the supply. While PACS (picture archiving and communication system) [3] has solved the retrieval and visualization part of the problem, a CAD (computer-aided diagnosis) system to generate diagnostic results from clinical MRI (magnetic resonance imaging) and CT (computed tomography) scans would not only reduce the burden on a radiologist, but also boost the confidence associated with a diagnosis. Occasionally, a CAD system, might also detect a disorder that a radiologist could have missed due to insufficient time to analyze a case. This realization motivates us to strive towards the development of a robust, accurate and fully automated system to detect lumbar

* Corresponding author. Tel.: +1 7164716651.
E-mail address: sghosh7@buffalo.edu (S. Ghosh).

http://dx.doi.org/10.1016/j.compmedimag.2014.03.005
0895-6111/

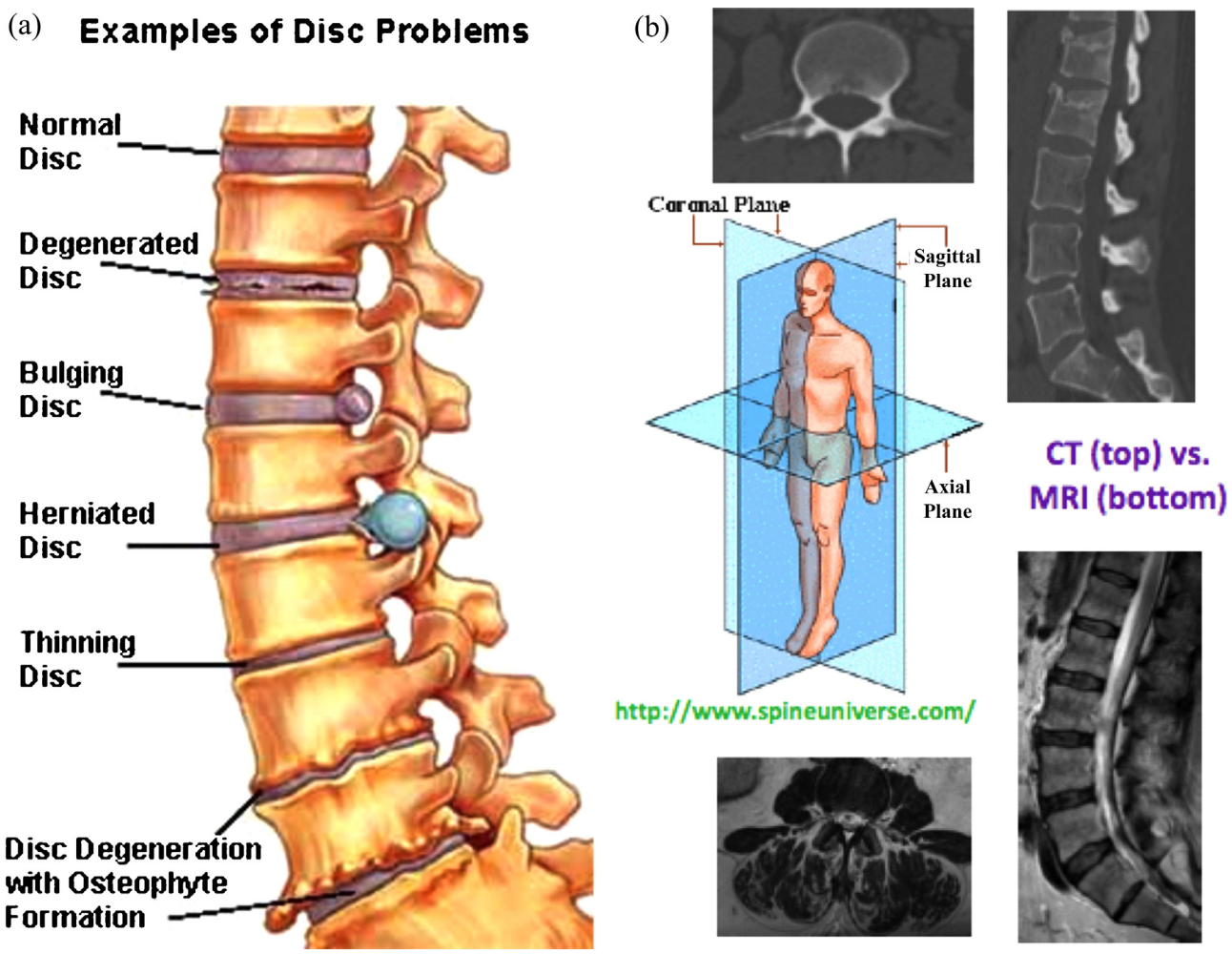

Fig. 1. Lumbar problems and current imaging techniques: (a) illustrates various disc problems in the lumbar spine while (b) illustrates the two common imaging modalities used for diagnosis.

abnormalities. MR and CT scans are two very popularly used modalities for diagnosis of lower back problems. While on one hand MRI is more expensive, it is non-invasive and better in terms of soft tissue detailing. Hence, it is a preferred modality for the diagnosis of intervertebral disc abnormalities like herniation, desiccation and degeneration. CT, on the contrary, is better suited for imaging boney structures. However, it uses harmful ionizing radiation leading to undesirable side-effects. Fig. 1(b) illustrates the MRI and CT imaging modalities for the lumbar spine along with the three imaging planes – sagittal, axial and coronal. In this paper, we focus only on clinical lumbar MRI, rather than full 3D volumes, since in everyday clinical routine, radiologists order separate spinal areas depending upon symptoms due to both cost and acquisition time-related issues.

Requirements for CAD systems of the spine are unique since we need to localize and correctly identify each intervertebral disc (i.e. label them as L5-S1, L4-L5 and so on), before we can proceed to the important task of detecting abnormalities. Localization of lumbar discs is a challenging problem due to a wide range of variabilities in the size, shape, count and appearance of discs and vertebrae. To this end, we first propose a robust method for labeling and localization of intervertebral discs in sagittal lumbar MRI images using machine learning methods and heuristics. This results in a tight rectangular bounding box for each lumbar disc which can be directly used for abnormality detection by extracting relevant features from disc bounding boxes as detailed in our earlier work [4,5]. Next, we propose a fully automatic approach to simultaneously segment the dural sac, discs and vertebra from clinical sagittal MRI using the neighborhood information of each pixel in an auto-context model.

In the subsequent sections, we discuss in detail past research (Section 2), the clinical dataset used in our experiments (Section 3), our approach for disc detection (Section 4) and disc segmentation (Section 6), along with detailed experimental results (Sections 5 and 7). Finally we draw our conclusion and discuss our future scope and research direction in Section 8.

2. Background and related work

Automatic detection of abnormalities from lumbar MRI scans has been studied by researchers for quite some time. The challenges are manifold – ranging from variations in scanner specifications, parameter settings, modalities, differences in body structure and composition, and last but not the least, the task of segmentation which is a major challenge in computer vision. Moreover, segmentation of MRI scans is quite difficult due to partial volume effect (where multiple tissues contribute to pixels and blurs intensity across boundaries), intensity inhomogeneities (non-anatomic intensity variations of the same tissue over the image due to RF non-uniformity, static field inhomogeneity, patient movement, etc.) and inter-organ/inter-tissue similarities (where two or more tissues/organs have similar gray-scale intensities).

2.1. Localization of lumbar tissues

There has been quite some research in the direction of automatic dural sac segmentation [6–8] and labeling and localization of intervertebral discs [9–11] from lumbar MRI.

Schmidt et al. [9] introduced a probabilistic inference method using a part-based model that measures the possible locations of the intervertebral discs in full back MRI. They achieve upto 97% part detection rate on 30 cases. Bhole et al. [12] presented a method for automatic detection of lumbar vertebrae and discs from clinical MRI by combining tissue property and geometric information from T1W sagittal, T2W sagittal and T2W axial modalities. They achieve 98.8% accuracy for disc labeling on 67 sagittal images. Alomari et al. [10] proposed a two-level probabilistic model that captures both pixel- and object-level features to localize discs. The authors use generalized EM (Expectation Maximization) attaining an accuracy of 89.1% on 50 test cases. Oktay et al. [13] proposed another approach using PHOG (pyramidal histogram of oriented gradients) based SVM and a probabilistic graphical model and achieved 95% accuracy on 40 cases. In all these works, the authors have concentrated on finding a point inside the disc, which immediately leads to the added requirement of a challenging segmentation step in order to diagnose a disc abnormality. Recently, we presented an approach [11] using heuristics and machine learning methods to provide tight bounding boxes for each disc achieving 99% localization accuracy on 53 cases. This method can by-pass complicated segmentation algorithms and directly feed the detected disc region to a CAD system that extracts relevant features and automatically provides diagnostic results [4,5]. The first part of this paper is an extension of our previous work [11] providing disc labeling and detection solutions for clinical lumbar MRI.

In the direction of spinal segmentation, Horsfield et al. [6] proposed a semi-automatic method for the segmentation of the spinal cord from MRI utilizing an active surface model to assess multiple sclerosis. Koh et al. [7] proposed an unsupervised and fully automatic method based on an attention model and an active contour model, achieving 0.71 Dice Similarity Index on 60 cases. Chen et al. [8] used a deformable atlas-based registration combined with a topology preserving classification to robustly segment the spinal cord and the CSF.

In most of the previous work, other than our recent attempt [14], segmentation of the dural sac and the intervertebral discs have been handled separately which might lead to overlapping tissue regions. Moreover, some techniques depend on shape models giving rise to errors in case of high variability in appearance. In our previous work [14], we used a Gibbs sampling approach to simultaneously label all tissues in the lumbar MRI. This method uses both neighborhood intensity information and label information for each update. Experimental results on 53 cases showed an average Similarity Index of 0.77 and 0.66 for the dural sac and intervertebral discs, respectively.

2.2. Diagnosis of lumbar abnormalities

There has been a growing interest in the research community for automatic diagnosis of lumbar abnormalities from MRI and CT scans. In a relatively early work, Tsai et al. [15] describe the detection of disc herniation from 3D MRI and CT volumes by using geometric features like shape, size and location. However, it is a computationally expensive method and serves better for visualization. Michopoulou et al. [16] showcased the classification of intervertebral discs into normal or degenerated, by using fuzzy-c means to perform semi-automatic atlas-based disc segmentation and then used a Bayesian clsssifier. They achieved 86–88% accuracy on 34 cases. They also reported 94% accuracy using texture features [17] for 50 manually segmented discs. Alomari et al. [18] presented a fully automated herniation detection system using GVF snake for an initial disc contour and then trained a Bayesian classifier on the resulting shape features. They achieved 92.5% accuracy on 65 clinical MRI cases but a low sensitivity of 86.4%. The same group also presented a desiccation diagnosis system in lumbar discs from clinical MRI [19] using a probabilistic model and achieving over 96% accuracy. These methods use shape and intensity features and have a high false negative rate. In our previous research [4,5], we have shown the design of a clinical herniation detection system along with comprehensive comparison of features, dimensionality reduction techniques and classifiers. However, this work has been validated on a very small dataset of 35 cases.

3. Our clinical lumbar MRI dataset

Clinical lumbar MRI used by our group is procured using a 3T Philips MRI scanner at Proscan Imaging Inc. It consists of manually co-registered T2 and T1 weighted sagittal views and T2 weighted axial views. Our detection and segmentation approaches are evaluated on two datasets: Dataset #1 (86 cases) was acquired before 2009; and Dataset #2 (126 cases) was acquired between 2010 and 2012.

T2 SPIR (selective partial inversion recover), is a special protocol which suppresses fat and, hence, shows good contrast between relevant lumbar tissues. It is very commonly used along with T2 weighted protocol for diagnosis, although due to time and cost factors SPIR scans might not be available. While all the cases in Dataset #1 also have T2 SPIR modality images, 114 out of 126 cases in Dataset #2 have this protocol.

All these cases were randomly picked, and they all have one or more abnormalities ranging from bulging discs, herniation, desiccation, degeneration, mild to severe stenosis, etc. For our experiments we use T2 weighted (or T2 SPIR) mid-sagittal slice for each case. We use our own labeling tool for manual segmentation, which performs B-spline interpolation to interactively give a smooth outline of segmented regions. We obtain manual disc segmentation, labeling all the visible discs (starting from L5-S1 at the bottom), the dural sac and the vertebral bodies.

4. Disc detection and labeling: our approach

In this section, we describe in detail our approach towards providing tight bounding boxes for all visible discs (starting from L5-L1 and upwards), specifically in clinical lumbar MRIs. Observing the clinical scans, we see that the technician acquires 6 axial slices for 4 or 5 lumbar intervertebral discs, changing the angle according to the orientation of the disc (Fig. 2(a)). Depending upon the case, axial views of 4 or 5 discs are recorded starting from L5-S1 and ending in either L2-L3 or L1-L2 giving rise to 24 or 30 axial slices.

We first localize the discs that have corresponding axial MRI by utilizing an approximate disc region calculated from the intersection of the axial slices with the sagittal as described in Section 4.2. Then we localize the remaining discs using a two stage classifier as detailed in Section 4.3. In both the steps we use the HOG features as described in Section 4.1. From here on we will refer to the discs with corresponding axial slices as the 'lower discs' and the rest of the discs as the 'upper discs'.

4.1. HOG feature computation

Histogram of Oriented Gradients (HOG) are feature descriptors used in computer vision and image processing for the purpose of object detection [20]. This technique counts occurrences of gradient orientation in localized portions of an image. For our experiments, given a sub-image, we divide it into $3 \times 3 = 9$ sub-windows and fix

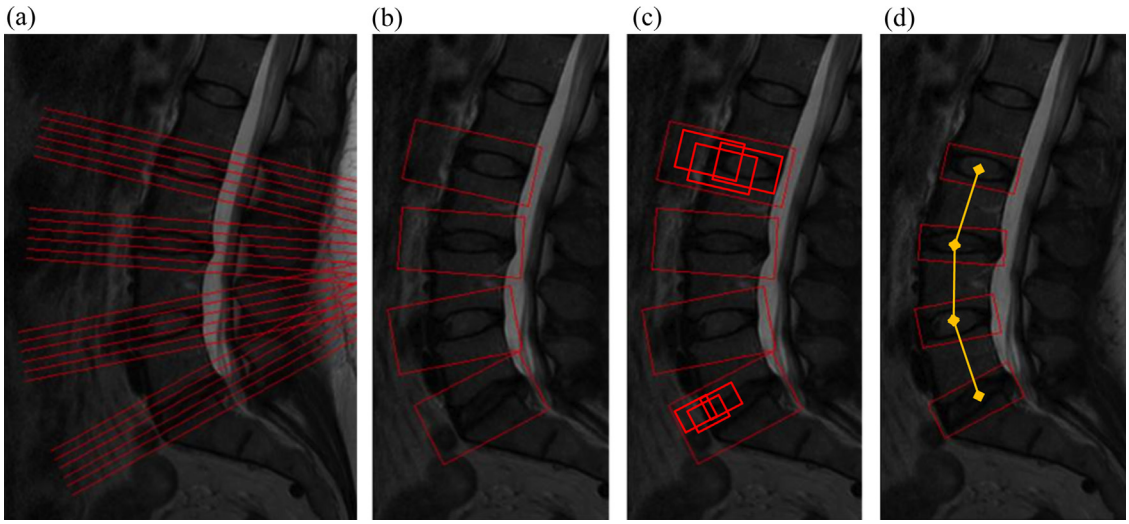

Fig. 2. Localization of the lower discs: (a) shows the lines of intersection of the mid-sagittal slice with all the available axial slices, (b) shows the rough lower disc ROIs extracted, (c) shows the candidate disc rectangles and (d) shows the final bounding box for each lower disc.

the bin size to 9. Thus our HOG feature is a vector of length 81, which is the only feature we use for disc prediction.

4.2. Localization of lower discs

1. *Extraction of rough bounding box*: We first extract a rough inclined rectangular bounding box for each lower disc using the lines of intersection (Fig. 1(a)) of the axial slices with the mid-sagittal slice and the high intensity spinal cord pixels to get an ROI as illustrated in Fig. 2(b).
2. *Creating the disc and non-disc training set*: For training we use T2 weighted mid-sagittal images of 50 cases which do not intersect with Dataset #1 and #2. From the manual disc labels, we create lower disc images using the inclination of the axial slices. We also create thousands of 60×20 non-disc images by sliding throughout the sagittal image.
3. *SVM Training*: We calculate HOG features for all the training images and model a binary SVMs [21], where the two classes are disc and non-disc. We use a linear kernel and fix the best parameters by 5-fold cross validation within the training set.
4. *Inclined rectangular tight bounding box*: After obtaining the rough disc regions for each lower disc (Fig. 2(b)), we extract thousands of multi-scale and multi-aspect ratio rectangles by sliding throughout the region. Then we calculate HOG features for these rectangles which contribute to the test set for the lower discs. Using the SVM modeled, we detect top candidate disc rectangles for each disc, and combine them by a weighted average to get the final tight bounding box as illustrated in Fig. 2(d). We also calculate the inter-disc-distance (*IDD*) of the lower discs, using the automatically detected disc centers in this step.

4.3. Localization of upper discs

1. *SVM training*: Similar to the lower discs, we create disc images using the manual labels of the upper discs, but this time the bounding boxes are not inclined, since we do not have axial slice information for these discs. We also create thousands of 80×40 non-disc images by sliding throughout the upper part of sagittal image (*IsagUpper*) decided by the upper-most axial line shown Fig. 3. We train a binary SVM using HOG features from the upper disc training images.
2. *Extraction of rough bounding box (Stage 1 classifier)*: We extract multi-scale and multi-aspect ratio rectangles by sliding throughout IsagUpper and calculate corresponding HOG features. Using the trained SVM model, we identify the disc rectangles (Fig. 3(a)) and heuristically remove outliers. We calculate the total number of upper visible discs (N_{upper}) from inter-disc distance (*IDD*) of the lower discs ($N_{upper} = D_{max}/IDD$, refer to Fig. 3(a) for D_{max}) and then cluster the disc rectangles into N_{upper} groups using k means. Next, we combine rectangles in each group to give a rough bounding box for each upper disc (Fig 3(c)).
3. *Rectangular tight bounding box (Stage 2 classifier)*: Finally we get a rectangular tight bounding box for each upper disc (Fig 3(d)) as we did for the lower discs.

Fig. 4 illustrates the workflow of our method.

5. Experimental results: disc detection

In this section we describe the metrics used to evaluate our disc detection approach and present out results.

5.1. Metrics

To evaluate the performance of our disc detection approach we calculate two commonly used metrics, deviation and accuracy as described:

1. Deviation of disc centers (Dev_{eu}) is the euclidean distance (in mm) between the center of the automatically detected and that of the manual disc bounding box.
2. Accuracy (*Acc*) is the percentage of automatic disc centers which visually lie inside the disc.

We also devise our own metrics since our output is a tight bounding box and not just a point within the disc.

1. Deviation of Percent Disc pixels (*DPD*): We define *DPD* as the deviation of the percentage of pixels in the manual bounding box belonging to disc (M_{per}) from the percentage of pixels in the automatic bounding box belonging to disc (A_{per}). We tabulate M_{per}, A_{per} and *DPD* to evaluate the tightness of the bounding box. Mathematically,

$$M_{per} = \frac{DiscPix_{manual}}{Pix_{manual}} \times 100; \quad A_{per} = \frac{DiscPix_{auto}}{Pix_{auto}} \times 100 \qquad (1)$$

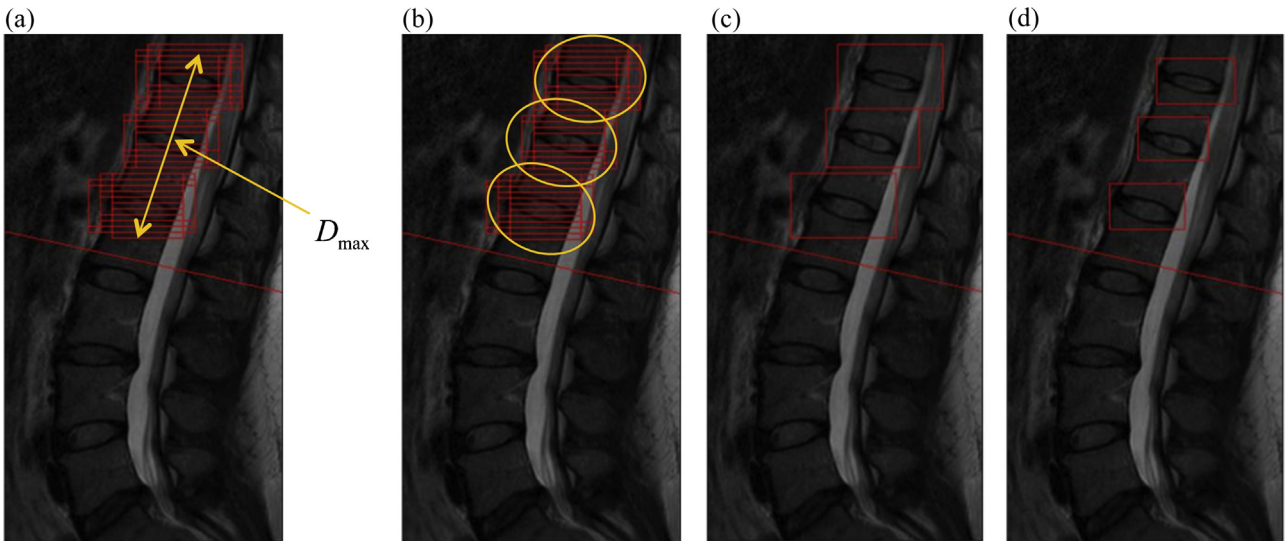

Fig. 3. Localization of upper discs – In (a), (b), (c) and (d), the inclined red line is the line of intersection with the uppermost axial slice. (a) Shows the candidate disc rectangles, (b) shows the N_{upper} disc clusters, (c) shows the rough bounding boxes for the upper discs and (d) shows the final automatically detected bounding boxes for all the upper discs.

where, $DiscPix_{manual}$ is the total number of disc pixels in the manual bounding box, Pix_{manual} is the total number of pixels in the manual bounding box, $DiscPix_{auto}$ is the total number of disc pixels in the automatic bounding box and Pix_{auto} is the total number of pixels in the automatic bounding box.

2. OutPercent (Out_{per}): Out_{per} is the percentage of disc pixels outside the automatic bounding box. It evaluates the accuracy of the bounding box.

$$Out_{per} = \frac{DiscPixOut_{auto}}{DiscPix_{manual}} \times 100 \quad (2)$$

where $DiscPixOut_{auto}$ is the total number of disc pixels outside the automatic bounding box.

5.2. Results and discussion

We evaluate our approach on our dataset, and calculate performance metrics using our manual segmentation, for both the T2 sagittal and SPIR images as shown in Tables 1 and 2. Note that the 'upper' discs may contain one or more thoracic intervertebral discs. Also note that the row 'Lumbar Avg' calculates the mean of the six lumbar disc metrics. Even though we detect all the visible discs starting from L5-S1 and upwards, we tabulate the performance results separately for only the lumbar discs, since they are the targeted ones in a lumbar MRI. We observe that the lower discs have tighter bounding boxes (lower *DPD*) than the upper ones, since we do not have corresponding axial information for the upper discs. The upper bounding boxes, being less tight, also have lower Out_{per}.

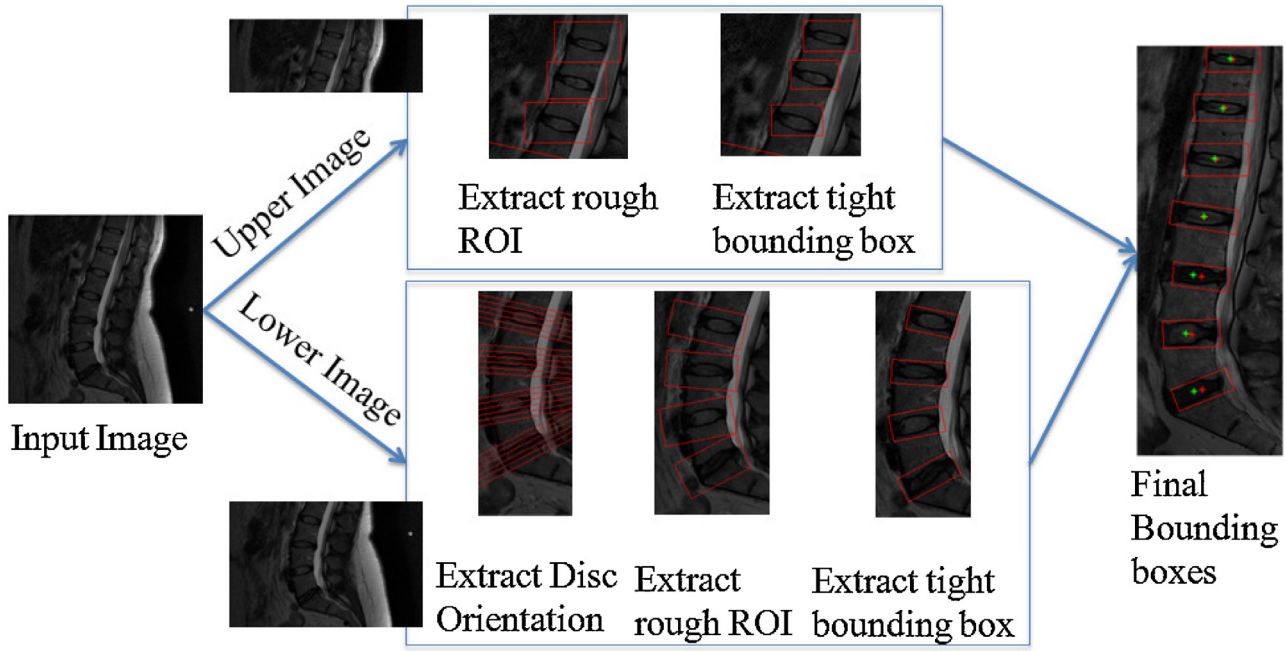

Fig. 4. This flow chart summarizes the steps in our approach for disc localization in lumbar MRI. The 'lower' discs are those which have corresponding axial slices, and the rest of the visible discs are the 'upper discs'. The extraction of a rough ROI and finally the tight bounding box for each disc is detailed in Sections 4.2 and 4.3.

Table 1
Automatic disc detection results on T2 sagittal data.

Disc label	Manual vs auto					
	M_{per} (%)	A_{per} (%)	DPD (%)	Out_{per} (%)	Dev_{eu} (mm)	Acc (%)
L5-S1	54.95	60.17	−5.22	6.18	2.82	93.40
L4-L5	60.04	63.67	−3.63	5.41	2.21	99.53
L3-L4	64.93	61.22	3.72	3.76	2.00	99.53
L2-L3	65.69	58.51	7.18	3.36	1.93	99.06
L1-L2	64.87	53.18	11.69	1.92	2.20	99.06
T12-L1	55.81	41.75	14.06	1.20	1.30	100.00
Lumbar Avg	61.05	56.43	4.63	3.64	2.08	98.43
Lower	62.05	59.45	2.60	4.16	2.23	98.10
Upper	55.61	39.23	16.38	1.39	1.41	97.33

Table 2
Automatic disc detection results on SPIR sagittal data.

Disc label	Manual vs auto					
	M_{per} (%)	A_{per} (%)	DPD (%)	Out_{per} (%)	Dev_{eu} (mm)	Acc (%)
L5-S1	55.17	57.45	−2.29	5.94	2.87	97.42
L4-L5	60.06	61.79	−1.73	5.15	2.60	100.00
L3-L4	65.05	59.71	5.34	3.39	2.32	99.48
L2-L3	65.66	56.98	8.68	3.43	2.43	98.45
L1-L2	64.94	52.81	12.13	2.95	2.74	98.97
T12-L1	55.86	41.64	14.22	1.58	2.08	98.96
Lumbar Avg	61.13	55.07	6.05	3.74	2.51	98.88
Lower	62.13	57.84	4.29	4.19	2.59	98.86
Upper	55.75	39.50	16.26	2.05	2.08	96.27

We achieve an average deviation of 2.08 mm for the T2 sagittal lumbar disc centers, which is better than the 3 mm average distance reported previously [13] for 40 cases. Moreover, our results are based on more than 200 clinical cases (effectively more than 400 images, since we use two modalities) that have a very wide range of variability. We achieve a disc detection accuracy of 98.43% for T2 and 98.88% for SPIR, which is the best reported so far. Unlike previous work [10,13], this method can also handle variable number of lumbar discs. Fig. 6 shows some representative samples of our disc detection approach which shows the effectiveness in a wide range of cases.

Probabilistic graphical models [10] usually take a long time to train and converge. Our method uses simple HOG features with linear SVM which makes disc detection faster. Also, with the advent of GPUs and frameworks like CUDA, features from sliding windows can be calculated in parallel, potentially giving high detection speeds. Currently, our sequential code in a PC takes 1 min. per case to detect all the visible discs. Our method outputs a tight bounding box for each disc instead of simply giving a point within [10,12,13]. Hence, we eliminate intermediate error-introducing segmentation steps and can directly feed the bounding box for relevant feature extraction and abnormality detection.

In the next section, we describe our approach to simultaneously segment the dural sac, intervertebral discs and vertebral discs which is an important precursor towards complete automatic quantification and localization of lumbar abnormalities.

6. Disc segmentation: our auto-context approach

In most of the previous work, other than our recent attempt [14], segmentation of the dural sac and the intervertebral discs have been handled separately which might lead to overlapping tissue regions. Moreover, some techniques depend on shape models giving rise to errors in case of high variability in appearance. Hence, we adopt a unified approach where we simultaneously label each pixel as belonging to one of four class labels (vertebra, intervertebral disc, dural sac or background) using the neighborhood information of each pixel in an auto-context model.

6.1. Auto-context model

Let us denote $\boldsymbol{X} = \{x_i : i \in \{1, 2\ldots, n\}\}$ as the set of pixel grayscale values in the mid-sagittal image. Our approach treats the

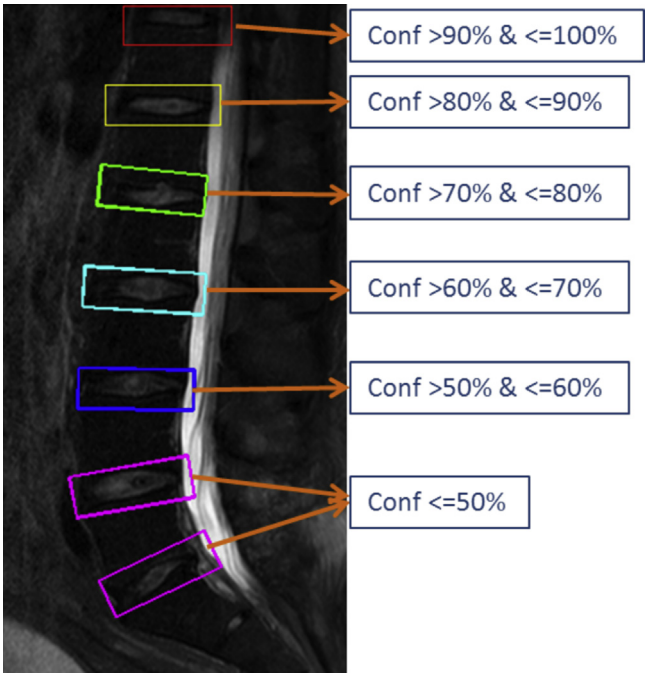

Fig. 5. This figure illustrates the color coding used to depict the confidence of the detected discs in Fig. 7. Note that this image does not show the actual confidence of the disc detected. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

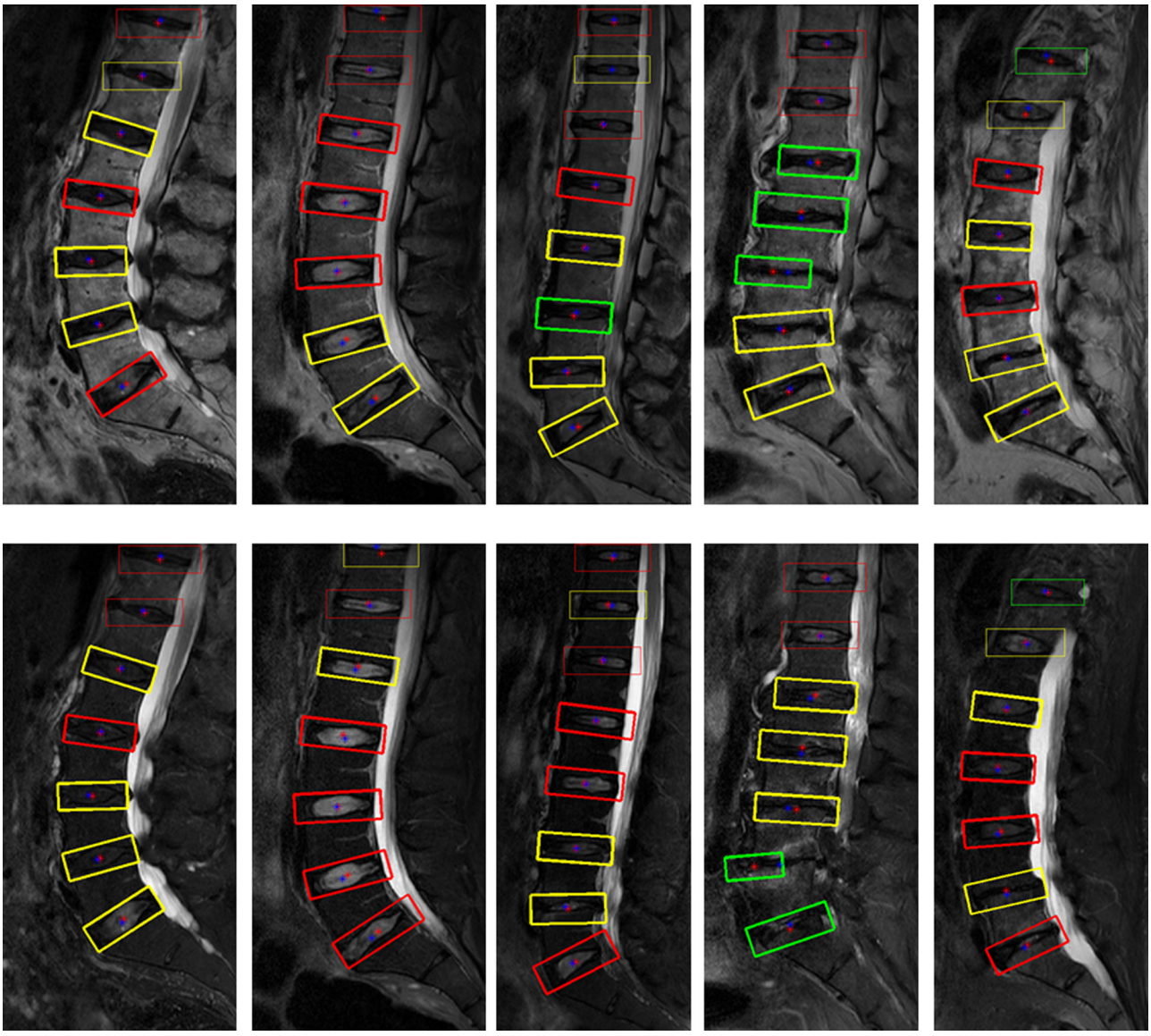

Fig. 6. This figure illustrates the automated disc detection results. The top row shows the results for some T2 sagittal images while the bottom shows results for the corresponding SPIR images. The red asterisk (*) are the automated disc centers while the blue ones are the manual centroids. The disc bounding boxes are color coded (refer to Fig. 5) according to a confidence measure derived from the probabilities of detected disc rectangles. The discs with the thin borders are the upper discs while the rest are lower discs. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

segmentation of lumbar MRI as a 4-class problem where each pixel can belong to any one of the following categories: vertebra, intervertebral disc, dural sac and background. The class labels are denoted by the set $\boldsymbol{L} = \{l : l \in \{1, 2, 3, 4\}\}$ and the set of pixel labels $\boldsymbol{Y} = \{y_i : i \in \{1, 2 \ldots, n\}, y_i \in \{\boldsymbol{L}\}\}$ where y_i is the output class label for the ith pixel.

Auto-context [22], is a general iterative learning framework used for segmentation which learns the low-level appearance, implicit shape, and context information through a sequence of discriminative models. This is accomplished by training a series of classifiers using the discriminative probability map of the previous classification, M^{t-1} where $M^t = (\boldsymbol{m}_1^t, \ldots, \boldsymbol{m}_n^t)$ at each time step t. Each vector $\boldsymbol{m}_i^t$ represents the probabilities of pixel i belonging to one of the k possible class labels, $\boldsymbol{m}_i = [P(y_i = 1), \ldots, p(y_i = k)]$. At time t, a classifier is trained to predict the true class label y_i given the features from image patch $nhood_i$ and the context information M_i^{t-1} where M is centered at pixel i. Once the classifier is trained, the new probability map M_t is used in the next iteration $(t+1)$ and the algorithm repeats until M converges. During testing, a new image has the same features extracted and goes through the iterative classification process using the learned probability distributions.

6.2. Implementation details

In our implementation, we use a patch size of 15×15 and calculate HOG (histogram of oriented gradient) vectors of length 81 as our low-level appearance features since they are proven to be robust for lumbar MRI [11]. The initial class probability map M^0 is set to have uniform values. We use an open-source implementation of random forests as classifier [23], empirically fixing the number of trees to 100. The total number of classifier stages is set to 5. For each iteration, we randomly sample 1500 pixels per training image per label, and train a classifier. Votes from the 100 trees are used to calculate the probability maps of neighboring pixels as context information at each iteration. We sparsely sample context pixels in 8 directions and 6 radii (5, 7, 16, 32, 64, 128) resulting in 192 context features.

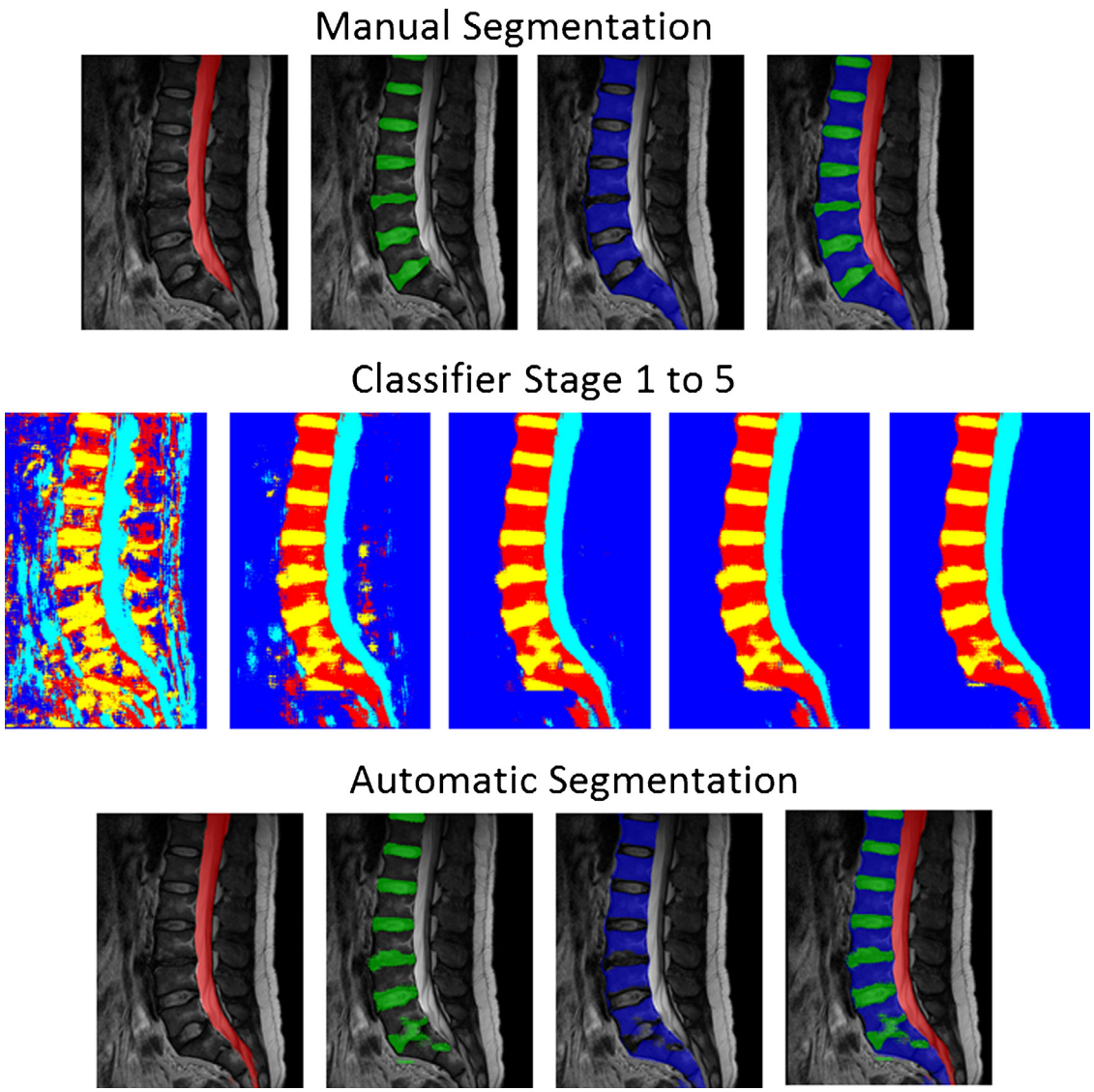

Fig. 7. This figure illustrates the segmentation results of one case. The first row shows the manual segmentation, the second shows the output of classifiers stages 1–5, and the third row shows the automatic segmentation results.

Fig. 7 shows an illustrative example of our segmentation method.

7. Experimental results: disc segmentation

In this section we describe in detail the experimental results of our auto-context approach for disc segmentation along with the metrics used for evaluation.

7.1. Metrics

Given the ground truth segmentation G and automatic segmentation M, we evaluate the validity of our approach using the following metrics:

1. Precision = $\frac{G \cap M}{M}$
2. Recall = $\frac{G \cap M}{G}$
3. Relative overlap (RO) = $\frac{G \cap M}{G \cup M}$
4. Similarity index (SI) = $\frac{2 \times (G \cap M)}{G + M}$

7.2. Results and discussion

Fig. 8 shows segmentation results of our auto-context method on a wide range of cases. The first row shows the manual segmentation, the second row shows the manual dural sac overlaid on the sagittal image to give a clear view of the discs and the third row shows the automatic segmentation results. Column 1 illustrates a highly deformed scenario where the L5-S1 segmentation fails completely. Column 2 shows a case with previous laminectomy at L4-L5, which results in failed automatic segmentation. Column 6 shows an instance where spinal stenosis causes the segmentation of the dural sac to get worse with each classifier stage, resulting in poor spinal segmentation.

The performance metrics from our experimental results are organized in Tables 3–5. Table 3 illustrates the overall performance of this method for both dural sac and intervertebral disc segmentation. T2 sagittal images show an average SI of 0.84 and 0.87 for disc and dural sac segmentation, respectively, which is far better compared to the previous work [14], that showed an average SI of 0.66 and 0.77 for the intervertebral discs and dural sac, respectively.

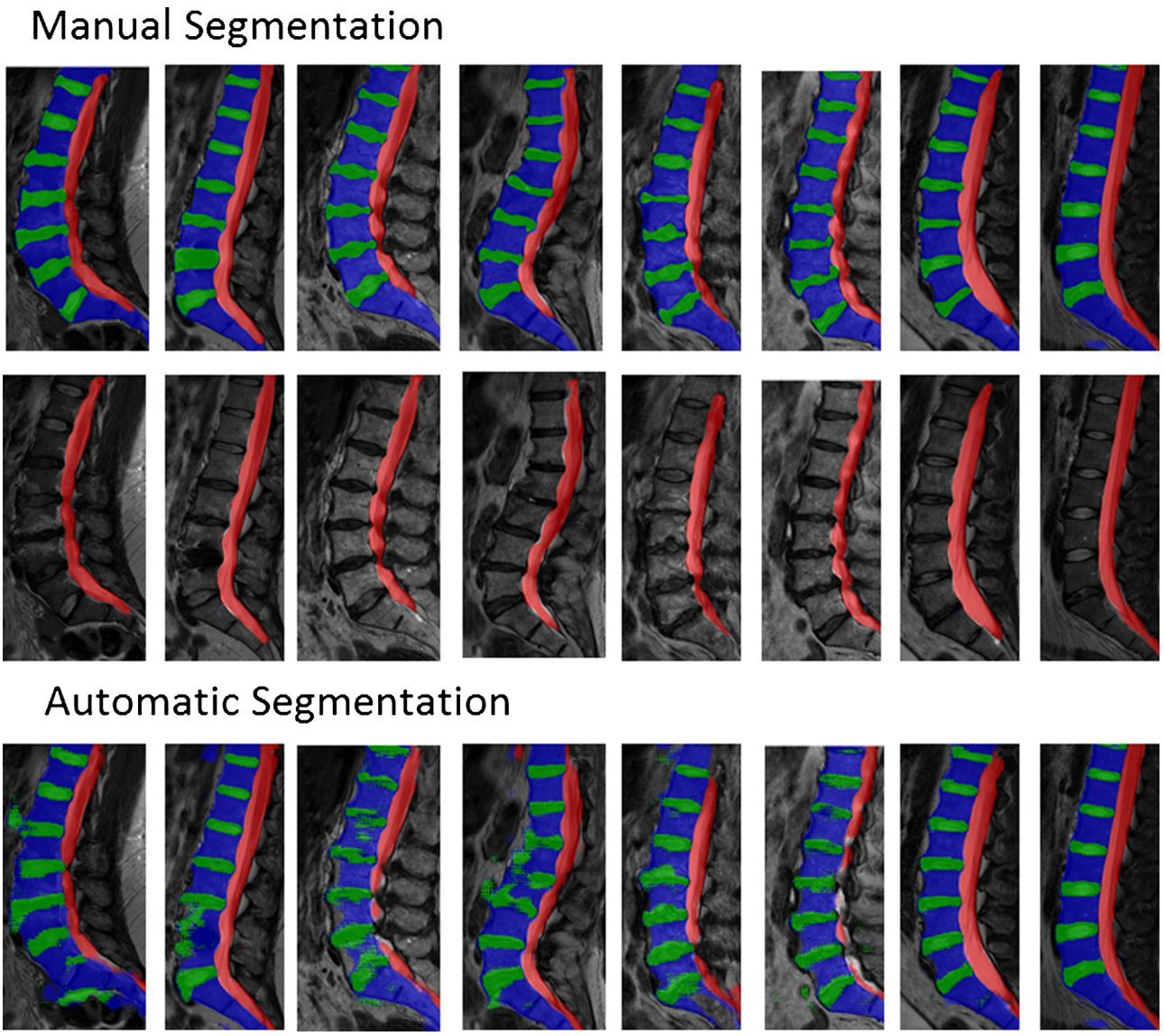

Fig. 8. Segmentation results of cases with a wide range of abnormalities.

Average Similarity Indices of 0.80 and 0.86 for SPIR disc and dural sac, respectively, shows that training on T2 sagittal images work on SPIR as well.

Tables 4 and 5 show the disc-wise segmentation performance of the T2 and SPIR images, respectively. Note that the first six discs (L5-S1 to T12-L1) are categorized as lumbar discs while the rest are thoracic discs that are sometimes visible in the lumbar scans. Also note that, due to abnormalities, all the lumbar discs may not be visible in the mid-sagittal scan. We observe from Tables 4 and 5 that L5-S1 discs perform poorly compared to the other lumbar discs. This is mainly due to the fact that L5-S1 discs are statistically more prone to degeneration. Moreover, they suffer from maximum variability not only when they are abnormal, but also when they are quite healthy. For instance, sometimes L5-S1 looks like a sacral disc due to sacralization of the L5 lumbar vertebra. Also, sometimes due to abnormal curvatures of the spine, the L5-S1 disc tends to get distorted in terms of inclination and shape.

Table 3
Overall results of the autocontext segmentation.

Metric	Dural-sac						Intervertebral disc					
	T2 SAG			T2 SPIR			T2 SAG			T2 SPIR		
	#1	#2	Avg	#1	#2	Avg	#1	#2	Avg	#1	#2	Avg
Precision	0.83	0.83	0.83	0.88	0.85	0.87	0.82	0.76	0.79	0.79	0.73	0.76
Recall	0.93	0.90	0.92	0.88	0.85	0.87	0.90	0.90	0.90	0.83	0.87	0.85
R O	0.79	0.76	0.78	0.78	0.74	0.76	0.76	0.70	0.73	0.70	0.66	0.68
S I	0.88	0.86	0.87	0.87	0.85	0.86	0.86	0.82	0.84	0.81	0.79	0.80

Table 4
Disc-wise segmentation results for T2 images.

Metric	Relative overlap			Similarity index		
	#1	#2	Avg	#1	#2	Avg
L5-S1	0.72	0.65	0.68	0.83	0.78	0.80
L4-L5	0.77	0.71	0.74	0.87	0.83	0.84
L3-L4	0.80	0.75	0.77	0.89	0.85	0.87
L2-L3	0.80	0.76	0.77	0.89	0.86	0.87
L1-L2	0.79	0.76	0.78	0.88	0.86	0.87
T12-L1	0.79	0.78	0.78	0.88	0.87	0.88
T11-T12	0.76	0.70	0.73	0.86	0.82	0.83
T10-T11	0.67	0.61	0.63	0.79	0.74	0.76

Table 5
Disc-wise segmentation results for SPIR images.

Metric	Relative overlap			Similarity index		
	#1	#2	Avg	#1	#2	Avg
L5-S1	0.62	0.57	0.59	0.74	0.71	0.72
L4-L5	0.72	0.68	0.70	0.83	0.80	0.81
L3-L4	0.76	0.72	0.74	0.86	0.83	0.84
L2-L3	0.75	0.72	0.73	0.85	0.83	0.84
L1-L2	0.77	0.73	0.75	0.87	0.84	0.85
T12-L1	0.76	0.74	0.75	0.86	0.85	0.85
T11-T12	0.70	0.68	0.69	0.81	0.80	0.80
T10-T11	0.71	0.67	0.68	0.82	0.80	0.80

To the best of our knowledge this method to simultaneously segment the discs, vertebrae and the dural sac from clinical lumbar MRI is the best reported so far, and brings us a little closer to automatic characterization of lumbar abnormalities.

8. Conclusion and future scope

In this paper, we proposed new supervised approaches towards detection and complete segmentation of clinical sagittal lumbar MRI.

First, we proposed a new approach towards intervertebral disc localization from lumbar MRI: one that effectively combines machine learning techniques with heuristics. We provide a novel end-to-end framework that outputs a tight bounding box for each disc, instead of simply marking the centroid of discs, as has been the trend in the recent past. In previous work, variations of probabilistic graphical models were used to detect disc centroids, and not only were they computationally expensive; they also suffered in terms of accuracy. Our approach, on the other hand, is not only simple and easy to implement, it is also computationally efficient and easily parallelizable. Moreover, our method targets detection and diagnosis from clinical 2D MRI, which is more popular in clinical practice compared to full 3D volumes due to lower acquisition time and cost. We have provided a novel solution to incorporate clinical acquisition heuristics into our method, making it more accurate and robust. Experiments on 212 clinical cases with a wide range of variabilities show encouraging results with a mean lumbar disc center deviation of 2.08 mm for T2 weighted sagittal images.

Second, we proposed a method to simultaneously segment all the tissues (vertebrae, intervertebral disc, dural sac and background) in a lumbar sagittal MRI using an auto-context approach instead of any explicit shape features or models. We train a series of robust classifiers (random forests) using image features (histogram of oriented gradients) and sparsely sampled context features, which implicitly represent the shape and configuration of the image. Past work tackles the lumbar segmentation problem on a tissue/organ basis, and which tend to perform poorly in clinical scans due to high variability in appearance. They also lead to overlapping tissue regions. On the other hand, we concentrate on a unified segmentation, which is inherently better since we take context into account. The only other work that simultaneously segments all lumbar tissues, is our previous work that uses a Gibbs Sampling approach for MAP estimation, using the pixel neighborhood appearance, a probability map and the neighborhood label information. Our approach is especially helpful for clinical lumbar MRIs, since, while on the one hand anatomical structures are mostly constrained to relatively fixed positions, on the other hand, the lumbar region can show extreme variability in the shape, structure and appearance of abnormal discs and the spine. This method performs substantially better than previously reported approaches, but there is still scope for improvement. It suffers from over-segmentation of the dural sac and intervertebral discs. In some cases, specially those with stenosis, the dural sac tends to become disconnected, which could potentially be solved using better context information like shape features. In the near future, keeping in mind our current encouraging results, we propose to validate the utility of this segmentation method towards robust diagnosis, localization and quantification of lumbar abnormalities like herniation and stenosis.

References

[1] Cherry DK, Hing E, Woodwell DA, Rechtsteiner EA. National ambulatory medical care survey: 2006 summary. Natl Health Stat Rep 2008;3:1–39.
[2] Bhargavan M, Sunshine JH, Schepps B. Too few radiologists? Am J Roentgenol 2002;178(5):1075–82.
[3] Long LR, Antani S, Lee D-J, Krainak DM, Thoma GR. Biomedical information from a national collection of spine x-rays: film to content-based retrieval. SPIE 2003;5033:70–84.
[4] Ghosh S, Alomari RS, Chaudhary V, Dhillon G. Computer-aided diagnosis for lumbar MRI using heterogeneous classifiers. In: Proceedings of the 8th IEEE international symposium on biomedical imaging: from nano to macro, ISBI. 2011. p. 1179–82.
[5] Ghosh S, Alomari RS, Chaudhary V, Dhillon G. Composite features for automatic diagnosis of intervertebral disc herniation from lumbar MRI. In: Proceedings of the 33rd annual international conference of the IEEE engineering in medicine and biology society, EMBC. 2011. p. 5068–71.
[6] Horsfield MA, Sala S, Neema M, Absinta M, Bakshi A, Sormani MP, Rocca MA, Bakshi R, Filippi M. Rapid semi-automatic segmentation of the spinal cord from magnetic resonance images: application in multiple sclerosis. Neuroimage 2010;50(2):446–55.
[7] Koh J, Scott PD, Chaudhary V, Dhillon G. An automatic segmentation method of the spinal canal from clinical MR images based on an attention model and an active contour model. In: Proceedings of the 8th IEEE international symposium on biomedical imaging: from nano to macro, ISBI. 2011. p. 1467–71.
[8] Chen M, Carass A, Oh J, Nair G, Pham DL, Reich DS, Prince JL. Automatic magnetic resonance spinal cord segmentation with topology constraints for variable fields of view. Neuroimage 2013;83:1051–62.
[9] Schmidt S, Kappes J, Bergtholdt M, Pekar V, Dries S, Bystrov D, Schnoerr C. Spine detection and labeling using a parts-based graphical model. In: IPMI'07: Proceedings of the 20th international conference on information processing in medical imaging. 2007. p. 122–33.
[10] Alomari RS, Corso JJ, Chaudhary V. Labeling of lumbar discs using both pixel- and object-level features with a two-level probabilistic model. IEEE Trans Med Imaging 2011;30(1):1–10.
[11] Ghosh S, Malgireddy MR, Chaudhary V, Dhillon G. A new approach to automatic disc localization in clinical lumbar MRI: combining machine learning with heuristics. In: Proceedings of the 9th IEEE international symposium on biomedical imaging: from nano to macro, ISBI. 2012. p. 114–7.
[12] Bhole C, Kompali S, Chaudhary V. Context-sensitive labeling of spinal structure in MRI images. Proc SPIE 2009;7260, 72603P–72603P-9.
[13] Oktay AB, Akgul YS. Localization of the lumbar discs using machine learning and exact probabilistic inference. In: MICCAI (3). 2011.
[14] Ghosh S, Malgireddy MR, Chaudhary V, Dhillon G. A supervised approach towards segmentation of clinical MRI for automatic lumbar diagnosis. In: Medical image computing and computer-assisted intervention (MICCAI) workshop on computational methods and clinical applications for spine imaging (CSI). 2013 (to be published).
[15] Tsai M, Jou S, Hsieh M. A new method for lumbar herniated inter-vertebral disc diagnosis based on image analysis of transverse sections. Comput Med Imaging Graph 2002;26(6):369–80.
[16] Michopoulou S, Costaridou L, Panagiotopoulos E, Speller R, Panayiotakis G, Todd-Pokropek A. Atlas-based segmentation of degenerated lumbar intervertebral discs from MR images of the spine. IEEE Trans Biomed Eng 2009;56:2225–31.
[17] Michopoulou S, Boniatis I, Costaridou L, Cavouras D, Panagiotopoulos E, Panayiotakis G. Computer assisted characterization of cervical intervertebral disc degeneration in MRI. J Instrum 2009;4:287–93.

[18] Alomari RS, Corso JJ, Chaudhary V, Dhillon G. Toward a clinical lumbar CAD: herniation diagnosis. Int J Comput Assist Radiol Surg 2011;6(1):119–26.
[19] Alomari RS, Corso JJ, Chaudhary V, Dhillon G. Desiccation diagnosis in lumbar discs from clinical MRI with a probabilistic model. In: Proceedings of the IEEE international symposium on biomedical imaging (ISBI'09). 2009. p. 546–9.
[20] Dalal N, Triggs B. Histograms of oriented gradients for human detection. In: International conference on computer vision and pattern recognition, vol. 2. 2005. p. 886–93.
[21] Chang CC, Lin CJ. LIBSVM: a library for support vector machines. ACM Trans Intell Syst Technol 2011;2(3).
[22] Tu Z, Bai X. Auto-context and its application to high-level vision tasks and 3D brain image segmentation. IEEE Trans Pattern Anal Mach Intell 2010;32(10):1744–57.
[23] Jaiantilal A. Classification and regression by randomforest-matlab; 2009 http://code.google.com/p/randomforest-matlab