CrossMark

ORIGINAL ARTICLE

Spine labeling in MRI via regularized distribution matching

Seyed-Parsa Hojjat[1] · Ismail Ayed[2] · Gregory J. Garvin[3] ·
Kumaradevan Punithakumar[4]

Abstract

Purpose This study investigates an efficient (nearly real-time) two-stage spine labeling algorithm that removes the need for an external training while being applicable to different types of MRI data and acquisition protocols.
Methods Based solely on the image being labeled (i.e., we do not use training data), the first stage aims at detecting potential vertebra candidates following the optimization of a functional containing two terms: (i) a distribution-matching term that encodes contextual information about the vertebrae via a density model learned from a very simple user input, which amounts to a point (mouse click) on a predefined vertebra; and (ii) a regularization constraint, which penalizes isolated candidates in the solution. The second stage removes false positives and identifies all vertebrae and discs by optimizing a geometric constraint, which embeds generic anatomical information on the interconnections between neighboring structures. Based on generic knowledge, our geometric constraint does not require external training.
Results We performed quantitative evaluations of the algorithm over a data set of 90 mid-sagittal MRI images of the lumbar spine acquired from 45 different subjects. To assess the flexibility of the algorithm, we used both T1- and T2-weighted images for each subject. A total of 990 structures were automatically detected/labeled and compared to ground-truth annotations by an expert. On the T2-weighted data, we obtained an accuracy of 91.6% for the vertebrae and 89.2% for the discs. On the T1-weighted data, we obtained an accuracy of 90.7% for the vertebrae and 88.1% for the discs.
Conclusion Our algorithm removes the need for external training while being applicable to different types of MRI data and acquisition protocols. Based on the current testing data, a subject-specific model density and generic anatomical information, our method can achieve competitive performances when applied to T1- and T2-weighted MRI images.

Keywords Spine labelling · Distribution matching · Magnetic resonance imaging (MRI) · Regularization · Geometric constraints

Introduction

Precise detection and identification of individual spine structures, e.g., the vertebrae and inter-vertebral discs, provide anatomical benchmarks that facilitate the evaluation and reporting of frequent inter-vertebral disc deformities such as protrusion [1]. Performing spine labeling manually is tedious and time-consuming, more so when the number of images in a spine series is large as is the case in magnetic resonance imaging (MRI),[1] which is the best modality to assess disc disorders [1]. MRI scans depict soft-tissue structures, thereby allowing characterization/quantification of disc bulges and protrusions. Automating or semi-automating spine labeling in MRI can reduce significantly the amount of user

✉ Ismail Ayed
ismail.benayed@etsmtl.ca

1 Deparment of Medical Imaging, Sunnybrook Health Sciences Centre, 2075 Bayview Ave., Toronto, ON M4N 3M5, Canada

2 Ecole de Technologie Superieure (ETS), 1100 Rue Notre-Dame O, Montreal, QC H3C 1K3, Canada

3 Department of Medical Imaging, University of Western Ontario, 268 Grosvenor St, London, ON N6A 4V2, Canada

4 Department of Radiology and Diagnostic Imaging, University of Alberta, 3A5.039, 8440 - 112 St NW, Edmonton, AB, Canada

---

1 Typically, spine MRI studies contain more than 100 images.

inputs, allowing thorough, reproducible and fast diagnosis of various spine diseases. This computational problem has recently attracted significant research attention [2–14]. In addition to their usefulness in spine diagnosis, such anatomical labels yield a natural patient-specific coordinate system that can be very useful in (i) mapping radiology reports to the corresponding image segments, (ii) building semantic inspection tools, and (iii) facilitating significantly other difficult spine image processing tasks. Indeed, such labeling techniques can guide image registration and fusion [15], and provide priors/constraints for image segmentation, retrieval, as well as shape and population analysis. For instance, several recent spine segmentation methods rely on a given labeling as input [16], or add a labeling process as a part of a segmentation algorithm [3]. Unfortunately, and as acknowledged by several studies [7,8,10], automating spine labeling is a difficult computational problem for several reasons:

– The intensity appearances of the structures of interest change significantly with the modality and acquisition protocol;
– There are wide variations in the positions, sizes and shapes of the spine, vertebrae and discs;
– The number of visible vertebrae and discs can vary from one image to another; and
– It is difficult to distinguish between individual spine structures because of their repetitive nature and similar appearances/shapes.

Prior art

Most of the existing algorithms require intensive and time consuming training from a large, manually-labeled data set (e.g., [5,7–11]), with the results often being dependent on (i) the choice of a specific training set and (ii) the choice of modality, image type and acquisition protocol. The purpose of such training phase is to build priors on the shapes/appearances of spine structures and their interconnections. These priors are subsequently embedded within a classification or regression technique using, for instance, deep convolution networks [5], graphical models [7,13,17], adaptive-boosting learning [8,14], support vector machines [11], random forest regression [9], or generative probabilistic models [10], among others. Although they can yield exceptional performance in cases similar to the training set, training-based algorithms may have difficulty in capturing the substantial variations encountered in a clinical context. For instance, the existing excellent algorithms that were recently designed for CT (e.g., [7,12,13]) are not directly applicable to MRI [8], mainly because of

the significant differences in appearances between these two modalities. In MRI, the difficulties are compounded by the intensity similarities and weak edges between different structures in the same spine image, abundant noise and significant variations in appearances and resolutions from one image to another. Such large variability also results from numerous types of MRI sequences (e.g., T1-weighted, T2-weighted, PD, etc.) and acquisition protocols. For instance, an algorithm designed and trained on T2-weighted MR data (e.g., [10]) may not be applicable to T1-weighted MR images. Moreover, training-based geometric models may not properly capture unseen (or pathological) cases outside the range of geometric characteristics learned from the training set. For instance, in [10], the authors embedded training-based information on the geometric interactions between pairs of discs, enforcing the distance between neighboring discs to fall within a range learned a priori from the training set. Therefore, a pathological case, which does not necessarily conform to these learned distances, may not be labeled correctly.

Summary of contributions

In this study, we propose an efficient (nearly real-time) two-stage spine MRI labeling algorithm that removes the need for an external training while being applicable to different types of MRI data. Based solely on the current image data, the first stage aims at detecting potential vertebra candidates following the optimization of a segmentation functional containing two terms: (i) a distribution-matching term encoding contextual information about the vertebrae via a density model learned from a very simple user input, which amounts to a point (mouse click) on a predefined vertebra; and (ii) a regularization constraint, which penalizes isolated candidates in the solution. The second stage removes false positives and identifies all vertebra and discs by optimizing a geometric constraint, which embeds generic anatomical information on the interconnections between neighboring spine structures. Based on generic knowledge, our geometric constraint does not require external training. We performed quantitative evaluations of the algorithm over a data set of 90 mid-sagittal MRI images of the lumbar spine acquired from 45 different subjects. To assess the flexibility of the algorithm, we used both T1- and T2-weighted images for each subject. A total number of 990 structures were automatically detected/labeled and compared to ground-truth annotations by a radiology expert. On the T2-weighted data, we obtained an accuracy of 91.6% for the vertebrae and 89.2% for the discs. On the T1-weighted data, we obtained an accuracy of 90.7% for the vertebrae and 88.1% for the discs.

Formulation

Regularized distribution matching functional

Let $\mathbf{D} : \Omega \subset \mathbb{R}^2 \to \mathcal{F} \subset \mathbb{R}^k$ be a function defined over spatial image domain Ω, and whose values are in a finite set $\mathcal{F}$ containing k-dimensional features (or bins). For each point $\mathbf{x} \in \Omega$, we build a feature vector $\mathbf{D}(\mathbf{x}) = [D_1(\mathbf{x}) \dots D_k(\mathbf{x})]$ that contains image statistics within several box-shaped image patches of different orientations/scales. Such patch-based features can encode contextual information about the vertebrae and their neighboring structures, e.g., size, shape, orientation and geometric relationships with neighboring structures. The purpose of this stage of our method is to detect potential vertebra candidates via a regularized distribution matching approach. Let $u : \Omega \to \{0, 1\}$ denotes a binary pixel classifier function, which defines a variable partition of Ω: $\{\mathbf{x} \in \Omega / u(\mathbf{x}) = 1\}$, corresponding to the classification (or segmentation) region containing the vertebra candidates, and its complement $\{\mathbf{x} \in \Omega / u(\mathbf{x}) = 0\}$, corresponding to non-vertebra pixels in Ω. We optimize the following functional with respect to u, which yields an initial two-region classification of the image domain:

$$E(u) = \underbrace{-\sum_{f \in \mathcal{F}} \sqrt{\mathbf{P}_u(f)\mathcal{M}(f)}}_{\text{Contextual Distribution matching}} + \underbrace{\lambda \int_{\Omega} \|\nabla u\| d\mathbf{x}}_{\text{Regularization}} \qquad (1)$$

In the following, we describe each of the variables and notations that appear in the regularized classification problem in (1):

- The regularization term is a standard total-variation (TV) penalty [18], which encourages the solution to be smooth. In the case of a discrete characteristic function u of a given set, the total variation of u is the perimeter of the set [18]. This term penalizes solutions with a large number of very small isolated regions. This facilitates inference in the next step, and keeps only strong vertebral candidates in the solution.
- $\mathbf{P}_u$ is the kernel density estimate (KDE) of the distribution of features $\mathbf{D}$ within the vertebral class (region) defined by $\{\mathbf{x} \in \Omega / u(\mathbf{x}) = 1\}$:

$$\mathbf{P}_u(f) = \frac{\int_{\Omega} u K(\mathbf{D} - f) d\mathbf{x}}{\int_{\Omega} u d\mathbf{x}} \qquad (2)$$

and K is the Gaussian kernel given by:

$$K(t) = \frac{1}{(2\pi\sigma^2)^{\frac{k}{2}}} \exp^{-\frac{\|t\|^2}{2\sigma^2}} \qquad (3)$$

with σ the width of the kernel.

- The first term in (1) is the negative Bhattacharyya coefficient, which evaluates the similarity (or affinity) between two distributions (or densities). The lower the value of this coefficient, the better the affinity between the distributions: the range of the distribution matching term is $[-1, 0]$, with -1 indicating a perfect match between the densities and 0 corresponding to a total mismatch. The Bhattacharyya coefficient is very popular in statistics for evaluating the affinity between densities, and it has the following geometric interpretation: it corresponds to the cosine of the angle between the *unit* vectors $(\sqrt{\mathbf{P}_u(f)}, f \in \mathcal{F})^T$ and $(\sqrt{\mathcal{M}(f)}, f \in \mathcal{F})^T$. Therefore, it considers explicitly $\mathbf{P}_u$ and $\mathcal{M}$ as distributions by representing them on the unit hypersphere.

  Minimization of this term aims at finding a classification region whose contextual-feature distribution most closely matches a model distribution $\mathcal{M}$. The latter can be either estimated interactively from a very simple user input placed on one of the images within a spine MRI series, or learned a priori from training data (i.e., a set of expert-labeled images different from the testing data set). In particular, our model is convenient in practice when we do not have a large set of relevant training images. As we will see later in the experiments, regularized classifier (1) does not require an intensive external learning of $\mathcal{M}$ to yield satisfying performances. We can estimate $\mathcal{M}$ from the feature vectors within a small region centered at a one-click user input (e.g., a point within a predefined vertebra such as L5).

- λ is a positive constant that balances the contribution of each term.

Optimization

Optimizing directly the high-order functional in (1) is a difficult problem because of (i) the high non-convexity (and non-linearity) of the distribution matching term in (1) and (ii) the integer-value constraint imposed on classification variable u. To obtain a solution efficiently, we split the problem into a sequence of sub-problems, each optimizing a bound (or auxiliary function) of (1) via a convex relaxation and the augmented Lagrangian method. An auxiliary function is an upper bound on the original energy, which verifies a certain condition. Optimizing an auxiliary function iteratively guarantees that the original energy decreases at each iteration ("Bound optimization and auxiliary function" section). The auxiliary function in this work is a sum of a *linear* function and a total-variation (TV) term. Therefore, unlike the original difficult functional, it can be optimized efficiently via a convex-relaxation (i.e., we relax the integer constraint) and an efficient multiplier-based algorithm. Specifically, for computing a global minimum of our relaxed auxiliary function, we solve an equivalent constrained prob-

lem via the standard augmented Lagrangian method [19–21], which embeds the constraint as a convex term ("Equivalent constrained problems and augmented Lagrangian method" section).

Bound optimization and auxiliary function

Bound optimization has recently led to competitive algorithms in the computer vision community, including efficient solutions to several difficult (high-order and integer-valued) segmentation functions, e.g., entropy [22], non-submodular pairwise terms [23] and distribution-matching constraints [24]. Rather than optimizing directly E, we optimize a sequence of auxiliary functions (bounds on E), denoted $A(u|u^i)$, $i \geq 1$ (i is the iteration number), whose optimization is easier than E:

$$u^{i+1} = \min_u \ A(u|u^i), \quad i \geq 1 \quad \text{s.t.} \tag{4a}$$

$$E(u) \leq A(u|u^i), \quad i \geq 1 \tag{4b}$$

$$E(u^i) = A(u^i|u^i) \tag{4c}$$

An auxiliary function of E is a function that verifies the constraints in (4b) and (4c). Using these constraints, and by definition of minimum in (4a), one can show that the sequence of solutions in (4a) yields a decreasing sequence of E:

$$E(u^i) = A(u^i, u^i) \geq A(u^{i+1}, u^i) \geq E(u^{i+1}) \tag{5}$$

Furthermore, $E(u^i)$ is lower bounded and, therefore, converges to a minimum of E. In this work, we consider the following bound on the negative Bhattacharyya coefficient, which we recently derived in [24].

Bound on the negative Bhattacharyya coefficient

Given a fixed u^i, for any $u : \Omega \to \{0, 1\}$ verifying $\{\mathbf{x} \in \Omega / u(\mathbf{x}) = 1\} \subset \{\mathbf{x} \in \Omega / u^i(\mathbf{x}) = 1\}$ and $\forall \alpha \in [0, \frac{1}{2}]$, we have the following bound (auxiliary function) on the negative Bhattacharyya coefficient [24]:

$$-\sum_{f \in \mathcal{F}} \sqrt{\mathbf{P}_u(f)\mathcal{M}(f)} \leq B_\alpha(u|u^i)$$
$$-\sum_{f \in \mathcal{F}} \sqrt{\mathbf{P}_{u^i}(f)\mathcal{M}(f)} = B_\alpha(u^i|u^i) \tag{6}$$

with B_α given by:

$$B_\alpha(u|u^i) = -(1+\alpha) \sum_{f \in \mathcal{F}} \sqrt{\mathbf{P}_{u^i}(f)\mathcal{M}(f)}$$
$$+ \int_\Omega \left\{ \alpha u g^i + (1+\alpha)(1-u)h^i \right\} d\mathbf{x}$$

where

$$g^i = \frac{\sum_{f \in \mathcal{F}} \sqrt{\mathbf{P}_{u^i}(f)\mathcal{M}(f)}}{\int_\Omega u^i d\mathbf{x}};$$

$$h^i(\mathbf{x}) = \frac{u^i(\mathbf{x})}{\int_\Omega u^i d\mathbf{x}} \sum_{f \in \mathcal{F}} K(\mathbf{D}(\mathbf{x}) - f) \sqrt{\frac{\mathcal{M}(f)}{\mathbf{P}_{u^i}(f)}} \tag{7}$$

It is easy to see that, using the result in (6), we have the following auxiliary function for our TV-regularized distribution matching functional in (1):

$$A_\alpha(u|u^i) = B_\alpha(u|u^i) + \lambda \int_\Omega \|\nabla u\| d\mathbf{x} \tag{8}$$

Convex relaxation of the auxiliary functions

The auxiliary function in (8) is the sum of a *linear* function B_α and a total-variation (TV) term. Therefore, it can be optimized efficiently via powerful convex-relaxation techniques, e.g., [20,21,25,26]. In fact, $\min_{u \in \{0,1\}} A_\alpha$ is still a non-convex problem due to the integer-valued constraint $u \in \{0, 1\}$. However, relaxing such constraint to interval $[0, 1]$ yields the following convex problem:

$$\min_{u \in [0,1]} A_\alpha(u|u^i) \tag{9}$$

Furthermore, we can use a standard and powerful result in the total-variation literature [18,20,21,25], which states that, for problems of the same form as our auxiliary function (i.e., a total-variation term combined with a linear term), simply thresholding the optimum $u^* \in [0, 1]$ of the relaxed convex problem yields a global optimum of the non-convex discrete problem over $u \in \{0, 1\}$.

Now, the following equivalence allows us to compute an exact and global optimum of A_α over $u \in \{0, 1\}$ using a multiplier-based augmented Lagrangian method [20,21,25].

Equivalent constrained problems and augmented Lagrangian method

The convex problem in (9) is equivalent to the following constrained problem (The proof follows the ideas of [21]):

$$\max_{p_h, p_g, p} \min_u \int_\Omega \left\{ p_h + u(\text{div } p - p_h + p_g) \right\} d\mathbf{x} \quad \text{s.t.}$$
$$p_h(\mathbf{x}) \leq (1+\alpha)h^i(\mathbf{x});$$
$$p_g(\mathbf{x}) \leq \alpha g^i(\mathbf{x}); \quad \text{and} \quad |p(\mathbf{x})| \leq \lambda \text{ a.e. } \mathbf{x} \in \Omega \tag{10}$$

where u is viewed as the multiplier to constraint div $p - p_h + p_g = 0$. $p : \Omega \to \mathbb{R}$, $p_h : \Omega \to \mathbb{R}$ and $p_g : \Omega \to \mathbb{R}$ are variables in the form of scalar functions. Furthermore, by simply thresholding the optimum $u^* \in [0, 1]$ of (10),

we obtain an exact and global optimum of the non-convex problem $min_{u \in \{0,1\}} A_\alpha(u|u^i)$. From the equivalence between (9) and (10), one can use an efficient multiplier-based algorithm for computing a local optimum of $E(u)$. The algorithm follows the standard augmented Lagrangian method [19–21]. This amounts to optimizing the following augmented Lagrangian function, which corresponds to our problem in (10) (Multiplier u is replaced here by v to avoid confusion between inner and outer iterations in the algorithm presented next):

$$L_c(p_h, p_g, p, v) = \int_\Omega \left\{ p_h + v(\operatorname{div} p - p_h + p_g) - \frac{c}{2} \right. $$
$$\left. \left\| \operatorname{div} p - p_h + p_g \right\|^2 \right\} d\mathbf{x} \qquad (11)$$

where c is a positive constant. A summary of the algorithm is given in Algorithm 1 (i and j are the numbers of inner end outer iterations respectively).

Algorithm 1: Multiplier-based augmented-Lagrangian optimization

- Initialize u by $u^0(\mathbf{x}) = 1$ ($i = 0$)
- **repeat**
  1. Update g^i and h^i according to the second line of (7)
  2. set $j = 0$, $p = p^0$, $p_h = p_h^0$, $p_g = p_g^0$ and $v = v^0$
  3. **repeat**
     - Optimize L_c with respect p:
       $p^{j+1} = \max_{|p| \leq \lambda} -\frac{c}{2} \left\| \operatorname{div} p - p_h^j + p_g^j - \frac{v^j}{c} \right\|^2$
       This can be solved by the Chambolle's algorithm [27].
     - Optimize L_c with respect to p_h (closed-form solution):
       $p_h^{j+1} = \max_{p_h < (1+\alpha)h^i} \int_\Omega$
       $\times \left\{ p_h - \frac{c}{2} \left\| \operatorname{div} p^{j+1} - p_h + p_g^j - \frac{v^j}{c} \right\|^2 \right\} d\mathbf{x}$
     - Optimize L_c with respect p_g (closed-form solution):
       $p_g^{j+1} = \max_{p_g \leq \alpha g^i} -\frac{c}{2} \left\| \operatorname{div} p^{j+1} \right.$
       $\left. -p_h^{j+1} + p_g - \frac{v^j}{c} \right\|^2$
     - Update multiplier v:
       $v^{j+1} = v^j - c(\operatorname{div} p^{j+1} - p_h^{j+1} + p_g^{j+1})$
     - Let $j = j + 1$
  until *Convergence*;
  4. Let $v*$ the solution obtained from the inner iterations above. Compute a binary solution by applying Otsu's thresholding to $v*$.
  5. Let $i = i + 1$
  6. Let u^i equal to the binary solution obtained at step 4.
 **until** *Convergence*;

The scheme is amenable to parallel implementations on graphics processing units (GPU), and yields nearly real-time solutions (<1 s) for labeling a typical 2D spine MRI

of size 512×512. This is an important advantage over graph-cut solutions for distribution matching [24], which do not accommodate parallel implementations. Also, it is worth noting that our optimization scheme can be extended to 3D. In our experiments, however, we focused on 2D evaluations because, from a clinical-application perspective, spine labeling in MRI is typically required on 2D mid-sagittal slices. One of the main applications is to use such mid-sagittal labels to facilitate navigation through axial slices via cross-reference features. Our 2D evaluations are consistent with previous works on labeling spine MRI, e.g., [10,28] (where the focus is also on 2D mid-sagittal slices).

Embedding high-level anatomic constraints

This second stage of our method removes the false positives obtained at the previous step, and identifies all the vertebrae and discs in the image. It consists of optimizing a geometric score (or constraint), which embeds anatomical information on the interconnections between neighboring spine structures. Based on anatomical knowledge, the proposed geometric score does not require any additional training effort. Let K denotes the number of connected regions obtained from the previous pixel-level classification step; see the examples in Fig. 4. Let $[z_1, \ldots, z_K] \in \Omega^K$ be the centroids of these regions, and y_{input} the user-provided point. This step consists of finding an optimal set of N points $[\hat{y}_1, \ldots, \hat{y}_N] \in \Omega^N$ among all the possible combinations of N points in $[z_1, \ldots, z_K]$ ($N < K$). We propose to optimize the following constraint, which, as illustrated in Fig. 1, embeds priors on the geometric relationships between spine structures:

$$[\hat{y}_1, \ldots, \hat{y}_N] = \max_{[y_1, \ldots, y_N] \subset [z_1, \ldots, z_K]} G(y_1, \ldots, y_N)$$
$$\text{s.t. } y_1 = y_{\text{input}} \quad \text{with}$$
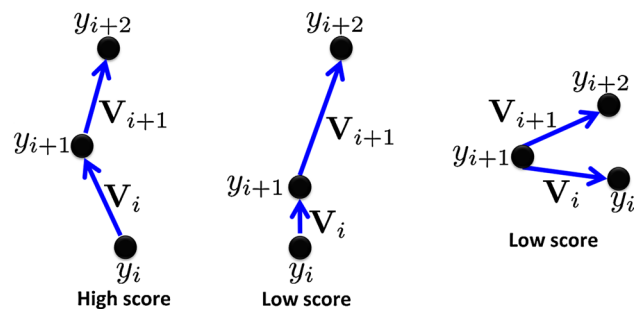
Fig. 1 Illustration of the geometric scores. *Left* plausible distances and angles between neighboring vertebrae; *middle* unlikely distance setting; *right* unlikely angle setting

$$G(y_1, \ldots, y_N) = \underbrace{\sum_{i=1}^{N-1} \cos(\mathbf{V}_i, \mathbf{V}_{i+1})}_{\text{Angles}}$$

$$+ \underbrace{\sum_{i=1}^{N-1} \min\left(\frac{\|\mathbf{V}_i\|}{\|\mathbf{V}_{i+1}\|}, \frac{\|\mathbf{V}_{i+1}\|}{\|\mathbf{V}_i\|} \right)}_{\text{Distances}} \qquad (12)$$

where $\mathbf{V}_i$ is the vector pointing from y_i to y_{i+1}, $i = 1, \ldots, N-1$: $\mathbf{V}_i = y_{i+1} - y_i$. The function in (12) enforces the following anatomical constraints (see Fig. 1 for an illustration):

– The distance constraint ensures that the distances between neighboring vertebrae are not significantly different. For instance, the geometry of the labeling in the middle of Fig. 1 is not likely: the distance between vertebrae (y_i, y_{i+1}) is much smaller than the distance between vertebrae (y_{i+1}, y_{i+2}).
– The angle constraint ensures that each three successive vertebrae correspond to a plausible (not excessive) bending of the spine. For instance, the geometry of the the three successive vertebrae in the right-hand side of Fig. 1 corresponds to a very sharp angle (or a high curvature), which is not consistent with the geometry of human spines (not even those showing significant deformities/abnormalities).

We find the optimum of (12) with an exhaustive search. We can vary N from 1 to K in the exhaustive search, and choose the combination that yields the highest value of G in Eq. (12). In our implementation, we start with a plausible minimum number of N, e.g., $N = 5$ (because, typically, a lumbar spine image contains at least the five vertebrae from L1 to L5). Then, we increase N gradually by searching for candidate points that yield the best possible increase of criterion G. For our problem, such an exhaustive search can be performed in real-time because our regularized distribution matching does not result in a very large number of connected regions (refer to the examples in Fig. 4). In our implementation, it took about $10^{-3}s$ for a typical 2D spine MRI scans of size 512×512. The optimal set $[\hat{y}_1, \ldots, \hat{y}_N]$ yields the target vertebrae. The target discs are computed from the optimum as follows: $t_i = \frac{\hat{y}_i + \hat{y}_{i+1}}{2} \ \forall i \in [1, \ldots, N-1]$ and $t_0 = \frac{3\hat{y}_1}{2} - \frac{\hat{y}_2}{2}$. The individual labels of the structures follow directly from the

detected structures by using the known spatial organization of the structures, e.g., L1 below T12, L2 below L1, etc.

Experiments

### Data set, model parameters and features

We performed quantitative evaluations of the algorithm over a data set of 90 mid-sagittal MRI images of the lumbar spine acquired from 45 different subjects (45 images of the type T2-weighted and 45 of the type T1-weighted). The purpose of using both the T1- and T2-weighted images of each subject is to assess the ability of the algorithm to handle different types of MRI data, with significant changes in appearances and various levels of resolution and noise. For the T1-weighted data, the slice thickness is in the range [3, 4 mm], and the pixel spacing is in the range [0.47, 0.57 mm]. For the T2-weighted data, the slice thickness is in the range [0.88, 4 mm], and the pixel spacing is in the range [0.41, 0.90 mm]. The details of the datasets are presented in Table 1. A total of number of 990 structures were automatically detected and labeled: in our evaluations, we focused on 11 lumbar structures/labels per image: 5 vertebra labels (T12, L1, L2, L3 and L4) and 6 disc labels (T12-L1, L1-L2, L2-L3, L3-L4, L4-L5 and L5-S1). Model $\mathcal{M}$ is estimated from the feature vectors within a disc centered at a very simple user input: a mouse click placed within the L5 vertebra for each subject. The parameters of the model were invariant for all the test images in the data set, and were fixed as follows: $\lambda = 1$ and $N = 6$. The contextual-feature function $\mathbf{D}$ is computed from the input image as follows. For each point $\mathbf{x} \in \Omega$, we built a feature vector of dimension 3: $\mathbf{D}(\mathbf{x}) = [D_1(\mathbf{x}), D_2(\mathbf{x}), D_3(\mathbf{x})]$, with D_1 the mean of intensity within a 3×10 rectangular-shaped, vertically-oriented patch, D_2 the mean of intensity within a 10×3 rectangular-shaped, horizontally-oriented patch, both centered at point $\mathbf{x}$, and D_3 the intensity of $\mathbf{x}$. Note that other features based on image gradients or textures can be used within the same framework. Finally, the radius of the disc for estimating model $\mathcal{M}$ is fixed equal to 5 pixels.

Typical examples and computational time

Figure 2 depicts typical results obtained with the proposed algorithm on T1-weighted and T2-weighted images with

Table 1 Details of the T1-weighted and T2-weighted MR datasets. Pixel resolution and slice thickness are given in mm

| Description	T2-weighted	T1-weighted
Image size	(320×320)–(640×640)	512×512
Pixel resolution	(0.41×0.41)–(0.90×0.90)	(0.47×0.47)–(0.57×0.57)
Slice thickness	0.88–4.00	3.00–4.00
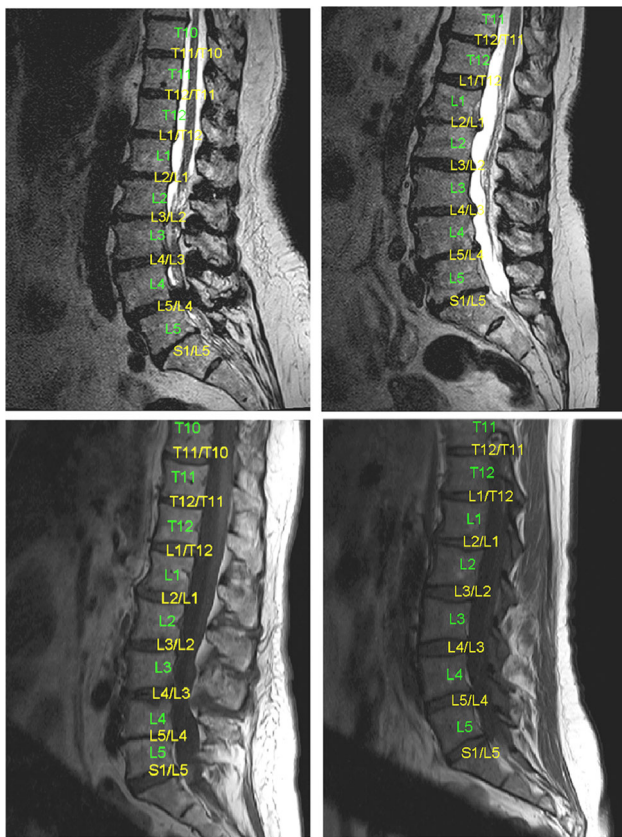
Fig. 2 Typical results obtained with the proposed algorithm on T1-weighted and T2-weighted images with significant changes in appearances and geometric characteristics such as the degree of bending of the spine and the shapes/sizes of individual vertebrae and discs. These examples show that the proposed algorithm handles effectively these variations, without the need for external training

significant changes in appearances and geometric characteristics such as the degree of bending of the spine and the shapes/sizes of individual vertebrae/discs. These examples show that the proposed algorithm effectively handles these variations, without the need for intensive external training. For each of these examples, the user click is placed at the L5 vertebra, and the solutions were obtained in near real-time: for a typical MRI scan of size 512×512, the first stage took less than $1s$, and the second took about $10^{-3}s$. Figure 3 depicts experiments with 5 different choices of the initial label. One can see from these typical examples that the choice of the initial label does not have a significant impact on the results. This makes sense because the feature distributions of the vertebrae are quite close. It is worth noting that from user perspective, L5 might be the easiest and fastest to identify. The convex relaxation optimizer was run in parallel on a graphics processing unit (GPU) of the type NVIDIA Quadro FX3700, with 112 Cuda cores.

Examples of intermediate segmentation results

Figure 4 depicts typical examples of the regularized distribution matching segmentation results we obtained using images from six different subjects (three T2-weighted in the first row of the figure and three T1-weighted in the third row). Below each image, we plot the corresponding evolution of the distribution matching term (the negative Bhattacharyya coefficient) as a function of the iteration number. Notice that, for each example, this segmentation step yielded a single connected region for each vertebral structure. Also, the number

Fig. 3 A typical T1-weighted MR example, which shows that the choice of the initial label (*red*) does not have a significant impact on the results
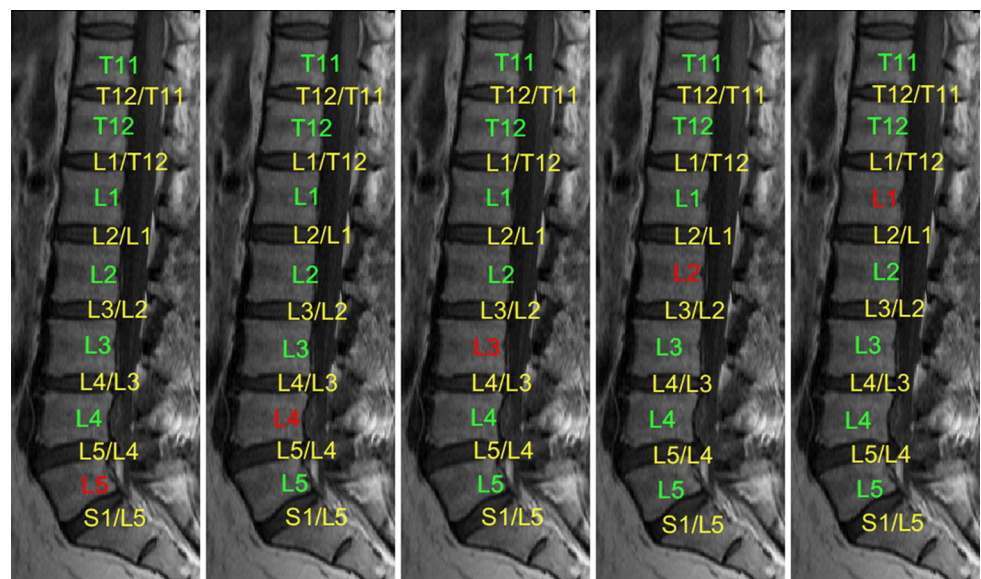
Fig. 4 Examples of regularized distribution matching and the corresponding evolutions of the distribution matching term as a function of the iteration number. *First and second row* T2-weighted examples; *third and fourth row* T1-weighted examples
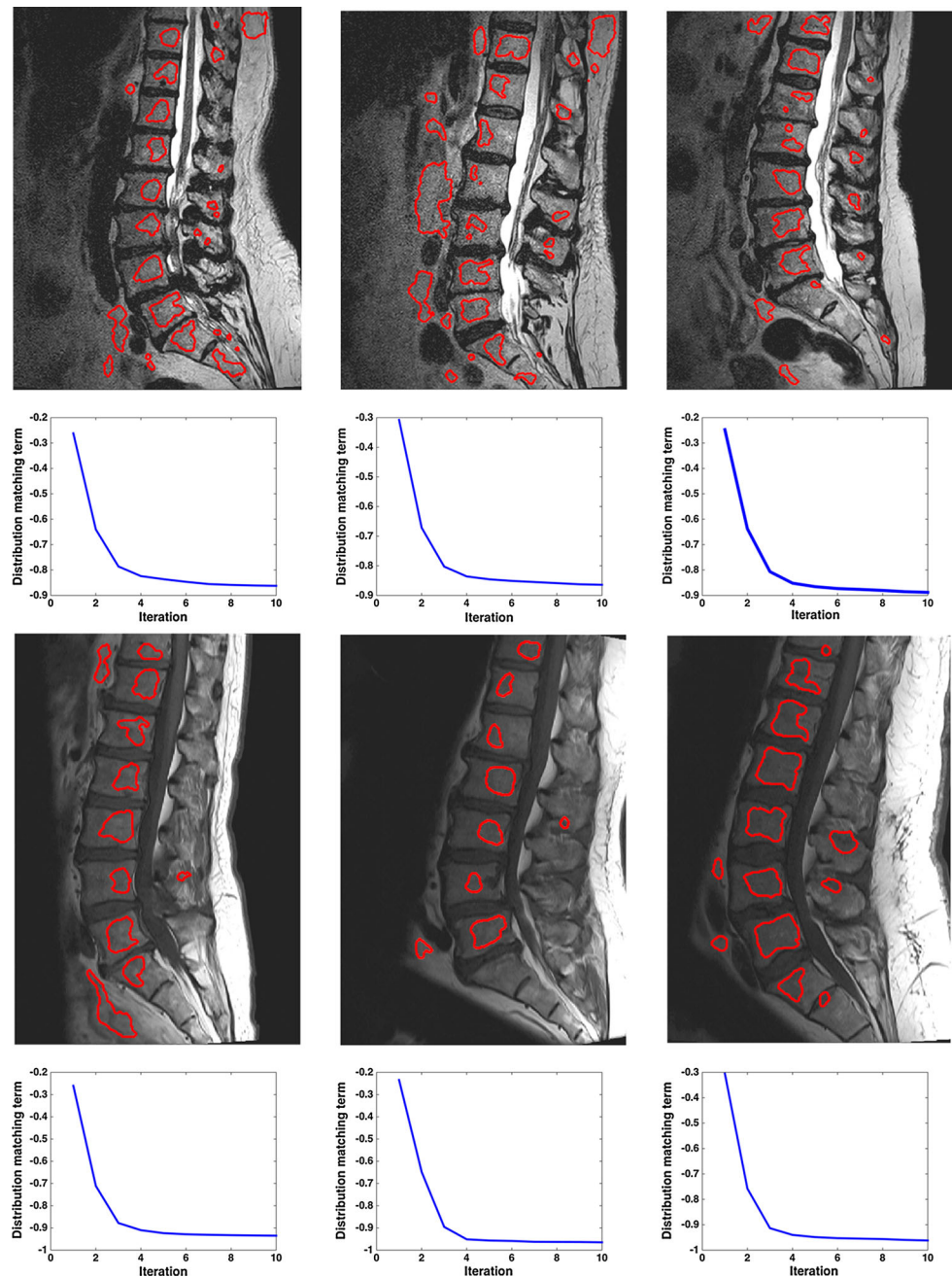
of the remaining irrelevant regions (i.e., the false positives that do not correspond to vertebral structures) is, typically, of the same order as the number of true positives (or even smaller in some cases). This shows how our regularized distribution matching discriminates well between the target structures and the rest of the image, significantly facilitating subsequent detection/annotation steps. Finally, it is worth noting that our distribution matching converges within a few iterations (typically less than 10 iterations); see the plots in Fig. 4.

Quantitative evaluations

We evaluated two performance measures: (i) the Euclidean distances (in mm) between the automatically obtained annotations and ground-truth annotations (Ground-truth labels are manually placed by a radiology expert at about the center of mass of each structure) and (ii) the detection accuracy given by the percentage of correctly detected and labeled vertebrae/discs. Similar to the study in [10], a structure (vertebra or disc) is considered correctly detected and
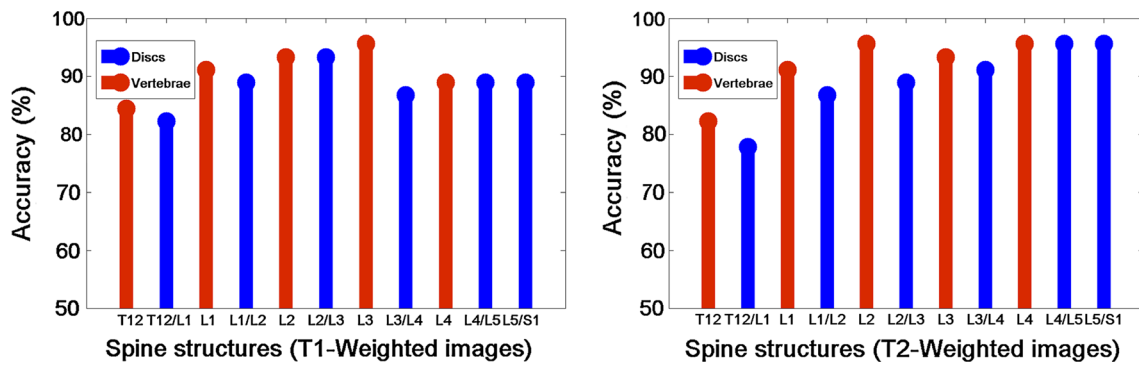
Fig. 5 Performance evaluations for each structure of the lumbar spine: percentage of correctly detected and labelled vertebrae/discs for the T1-weighted images (*left*) and T2-weighted images (*right*)
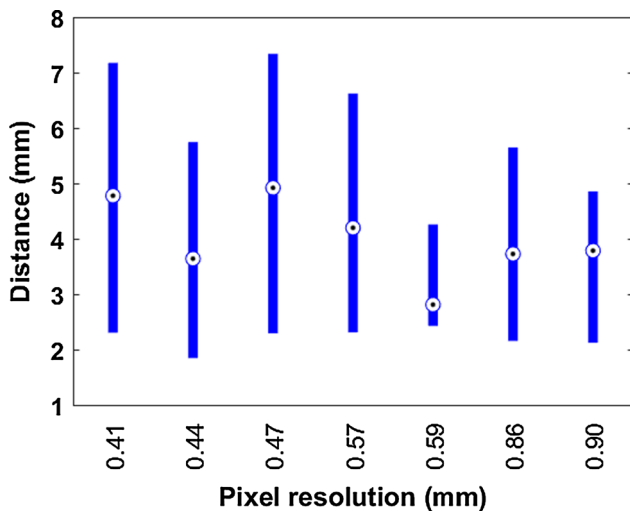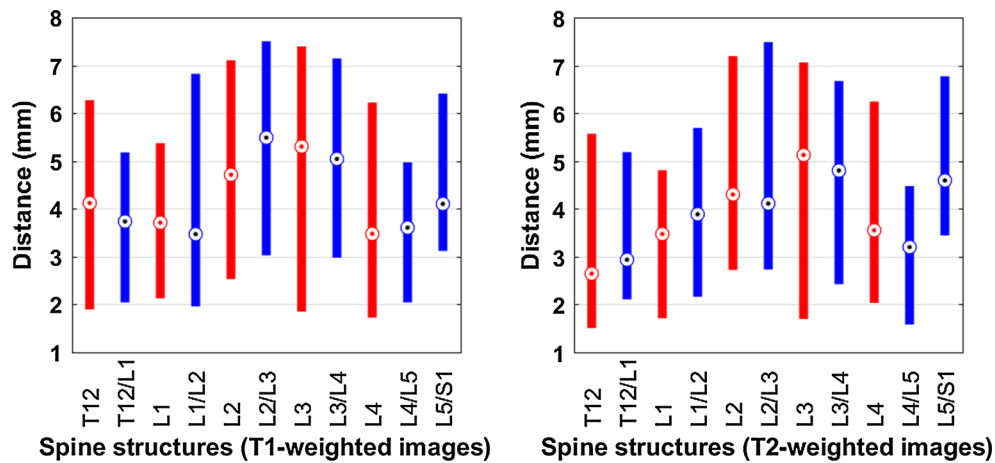

Fig. 6 The median, 5th percentile and 95th percentile values of Euclidean distances (in mm) between the automatically obtained annotations and ground-truth annotations for different pixel resolutions. Comparisons over 990 annotations showed that the performance of the algorithm did not vary significantly with pixel resolution

labelled if (a) the detected point is within the structure and (b) the automatic label is consistent with the ground-truth label. Table 3 reports the percentages of correctly detected and labeled vertebrae/discs over all the data set, and the Euclidean distances (expressed in mean ± standard deviation) between the automatically obtained annotations and ground-truth annotations. Figures 5 and 7 plot the detailed performances of our method for each individual structure of the lumbar spine. Figure 6 depicts the Euclidean distances (in mm) between the automatically obtained annotations and ground-truth annotations for the different pixel resolutions in our data set. This plot illustrates how the performance of the algorithm did not vary significantly with pixel spacing.

Comparison with prior work

Table 2 reports a meta-analysis of accuracy[2] for several state-of-the-art spine labeling techniques, including: the training-based method in [28], which proposed a deformable part model object detector combined with dynamic programming; the probabilistic model in [10], which integrates both pixel-level (appearance and shape) and context; the deep learning technique in [29], which combines image-to-image network, message passing and sparsity regularity; the method in [30], which uses supervised classification forests; and the deep convolutional neural network method in [31]. The performance of these techniques are reported in term of the percentage of correctly labeled structures. For each method, we report a range of accuracies as the authors of these studies report several performances, each corresponding to a different proportion of testing/training data, a different imaging protocol, or a differnt part of the spine (lumbar, thoracic or cervical). Notice that, for each of the methods (including ours), the best performance is close to 90% accuracy. Of course, we do not expect our algorithm to outperform training-based techniques because our algorithm uses much less information. However, the results (Table 3) suggest that the proposed algorithm can reach a performance comparable to training-based algorithms, although it does not use any external training data (i.e., model learning in our case is subject-specific and is based on a very simple user input). The study in [10] reported a performance comparable to us over an MRI data set of about the same size (105 subjects, with about half of the population for testing and the rest for training): the accuracy [10] is in the range [87.7, 89.1%]. However, the method in [10] is based on intensive training information from annotated data sets of over 50 subjects. It is worth

---

[2] Our meta-analysis of accuracy is not a direct comparison between the methods because the used data sets are different, and might have various levels of difficulties. For instance, the CT data set in [29] contains several unusual appearances (e.g., abnormal spine curvature), and undergo large variations in the field of view, image noise and resolutions.

Fig. 7 The median, 5th percentile and 95th percentile values of Euclidean distances (in mm) between the automatically obtained annotations and ground-truth annotations for the T1-weighted images (*left*) and the T2-weighted images (*right*)

Table 2 A meta-analysis of accuracies for several state-of-the-art spine labeling techniques

Method	Training	Accuracy range	Dataset
Alomari et al. [10]	Yes	[87.7, 89.1%]	105 MRIs (T2-SPIR weighted)
Lootus et al. [28]	Yes	[84.1, 92.9%]	371 MRIs (T2-weighted)
Yang et al. [29]	Yes	[78, 93%]	302 CT images
Glocker et al. [30]	Yes	[70, 89%]	302 CT images
Chen et al. [31]	Yes	[76, 92%]	302 CT images
Proposed algorithm	No	[88.1, 91.6%]	90 MRIs (T1- and T2-weighted)

Table 3 The overall labeling performances and Euclidean distances (in mm) with respect to ground-truth labels

Modality	Accuracy		Euclidean distances (mm)	
	Vertebra (%)	Disc (%)	Vertebra	Disc
T2-weighted	91.6	89.2	4.62 ± 3.19	4.54 ± 2.69
T1-weighted	90.7	88.1	5.12 ± 3.09	4.94 ± 2.88

noting that a systematic comparison between our interactive method and such training-based methods on the same data is not possible because our algorithm does not use any external training information. In fact, our method is not in direct competition with these learning-based techniques. It can be used when training information is not available, or as a tool that can efficiently build large annotated data sets for learning-based algorithms (e.g., convolutional neural networks)

Discussion

We proposed a fast (nearly real-time) spine MRI labeling algorithm that removes the need for external training while being applicable to different types of MRI data and acquisition protocols. The algorithm was based solely on the current testing image and a subject-specific model density, which is learned from a very simple user input. When applied to T1- and T2-weighted MRI images, our regularized distribution-matching formulation, followed by a simple geometric filtering step, achieved performances comparable

to state-of-the-art training-based techniques. Such comparisons are based on a meta-analysis of accuracies. In fact, our method is not in direct competition with these learning-based techniques: a systematic comparison on the same data (including training images) is not possible because our algorithm does not use any external training information. Our interactive method can be used when training information is not available or, also, to rapidly build massive annotated data sets, which can boost the performances of learning-based algorithms (e.g., convolutional neural networks).

The fact that our algorithm does not use external training information is a practical advantage over the existing methods. However, it comes at the price of a user input (a mouse click). Such an input makes the problem easier because off-by-one errors are less likely to occur. Also, learning a distribution from the testing image makes the algorithm flexible: it can be readily applied to different types of MRI data and acquisition protocols without re-training.

Our method was evaluated on patient data, but we did not focus on a specific pathology or injury. Our goal was to design

a system that performs well for the majority of patient data sets that are collected from clinical routine MRIs of the spine, with no exclusion criteria. Our data involved spine patients with typical disorders such as herniation or degenerative disc disease (DDD).

One of the limitations of the method is that the geometric refinement step does not account for missing structures. Another limitation is that the image features that we used for building the distribution-matching term are hand-crafted. There are several possible extensions to address these limitations. As our algorithm runs in nearly real-time, an interesting and practical extension would be to embed minimal user interactions in the form of constraints so as to rapidly correct possible errors in the results. Such an extension would prompt the user to add interactively new clicks, which the algorithm account for in a subsequent result-improvement step. For instance, interactive manual corrections would be useful when the method fails in labeling some parts of the spine, particularly in cases of atypical images such as those of patients with spine injuries. Another extension would be to learn the range of our geometric features (e.g., the angles) from a few annotated examples, or to learn a set of image features in order to build the distribution matching term. Also, there are several possible evaluations of the proposed method and its extensions, which we intend to investigate in the future. For instance, it would be interesting to evaluate the algorithm in the context of other modalities, e.g., computed tomography (CT), or on Thoracic and Cervical spine. We anticipate that our method works reasonably on these parts of the spine as well. Similarly to the lumbar spine, the Thoracic/Cervical features should not change significantly from one vertebra to another and the geometric prior seems to be plausible.

Compliance with ethical standards

Conflict of interest The author(s) declare that they have no competing interests.

Ethical approval The study was approved by the University of Western Ontario Research Ethics Board.

References

1. Fardon DF, Milette PC (2001) Nomenclature and classification of lumbar disc pathology: recommendations of the combined task forces of the North American Spine Society, American Society of Spine Radiology, and American Society of Neuroradiology. Spine 26(5):E93–E113
2. Wimmer M, Major D, Novikov AA, Buhler K (2016) Local entropy-optimized texture models for semi-automatic spine labeling in various MRI protocols. In: International symposium on biomedical imaging (ISBI), pp 155–159
3. De Leener B, Cohen-Adad J, Kadoury S (2015) Automatic segmentation of the spinal cord and spinal canal coupled with vertebral labeling. IEEE Trans Med Imaging 34(8):1705–1718
4. Ullmann E, Paquette JFP, Thong WE, Cohen-Adad J (2014) Automatic labeling of vertebral levels using a robust template-based approach. Int J Biomed Imaging 719520(1–719520):9
5. Cai Y, Landis M, Laidley DT, Kornecki A, Lum A, Li S (2016) Multi-modal vertebrae recognition using transformed deep convolution network. Comput Med Imaging Graphics 51:11–19
6. Kelm BM, Wels M, Zhou SK, Seifert S, Sühling M, Zheng Y, Comaniciu D (2013) Spine detection in CT and MR using iterated marginal space learning. Med Image Anal 17(8):1283–1292
7. Glocker B, Feulner J, Criminisi A, Haynor DR, Konukoglu E (2012) Automatic localization and identification of vertebrae in arbitrary field-of-view CT scans. Med Image Comput Comput Assist Interv MICCAI 3:590–598
8. Zhan Y, Dewan M, Harder M, Zhou X S (2012) Robust MR spine detection using hierarchical learning and local articulated model. Med Image Comput Comput Assist Interv MICCAI 1:141–148
9. Roberts MG, Cootes TF, Adams JE (2012) Automatic location of vertebrae on DXA images using random forest regression. Med Image Comput Comput Assist Interv MICCAI 3:361–368
10. Alomari RS, Corso JJ, Chaudhary V (2011) Labeling of lumbar discs using both pixel- and object-level features with a two-level probabilistic model. IEEE Trans Med Imaging 30(1):1–10
11. Oktay AB, Akgul YS (2011) Localization of the lumbar discs using machine learning and exact probabilistic inference. Med Image Comput Comput Assist Interv MICCAI 3:158–165
12. Ma J, Lu L, Zhan Y, Zhou XS, Salganicoff M, Krishnan A (2010) Hierarchical segmentation and identification of thoracic vertebra using learning-based edge detection and coarse-to-fine deformable model. Med Image Comput Comput Assist Interv MICCAI 1:19–27
13. Klinder T, Ostermann J, Ehm M, Franz A, Kneser R, Lorenz C (2009) Automated model-based vertebra detection, identification, and segmentation in ct images. Med Image Anal 13(3):471–482
14. Huang S, Chu Y, Lai S, Novak C (2009) Learning-based vertebra detection and iterative normalized-cut segmentation for spinal MRI. IEEE Trans Med Imaging 28(10):1595–1605
15. Miles B, Ayed I, Law MWK, Garvin GJ, Fenster A, Li S (2013) Spine image fusion via graph cuts. IEEE Trans Biomed Eng 60(7):1841–1850
16. Wang Z, Zhen X, Tay K, Osman S, Romano W, Li S (2015) Regression segmentation for m^3 spinal images. IEEE Trans Med Imaging 34(8):1640–1648
17. Schmidt S, Kappes JH, Bergtholdt M, Pekar V, Dries SPM, Bystrov D, Schnörr C (2007) Spine detection and labeling using a parts-based graphical model. In: Information processing in medical imaging (IPMI), pp 122–133
18. Nikolova M, Esedoglu S, Chan TF (2006) Algorithms for finding global minimizers of image segmentation and denoising models. SIAM J Appl Math 66(5):1632–1648
19. Bertsekas DP (1999) Nonlinear programming. Athena Scientific, Belmont
20. Yuan J, Bae E, Tai X-C, Boykov Y (2010) A study on continuous max-flow and min-cut approaches. Part I: binary labeling, technical report CAM-10-61, UCLA
21. Yuan J, Bae E, Tai X-C (2010) A study on continuous max-flow and min-cut approaches. In: IEEE conference on computer vision and pattern recognition (CVPR)
22. Tang M, Ben Ayed I, Boykov Y (2014) Pseudo-bound optimization for binary energies. Eur Conf Comput Vis ECCV 5:691–707

23. Taniai T, Matsushita Y, Naemura T (2015) Superdifferential cuts for binary energies. In: IEEE conference on computer vision and pattern recognition (CVPR), pp 2030–2038

24. Ben Ayed I, Punithakumar K, Li S (2015) Distribution matching with the bhattacharyya similarity: a bound optimization framework. IEEE Trans Pattern Anal Mach Intell 37(9):1777–1791

25. Punithakumar K, Yuan J, Ayed IB, Li S, Boykov Y (2012) A convex max-flow approach to distribution-based figure-ground separation. SIAM J Imaging Sci 5(4):1333–1354

26. Salah MB, Ben Ayed I, Yuan J, Zhang H (2014) Convex-relaxed kernel mapping for image segmentation. IEEE Trans Image Process 23(3):1143–1153

27. Chambolle A (2004) An algorithm for total variation minimization and applications. J Math Imaging Vis 20(1–2):89–97

28. Lootus M, Kadir T, Zisserman A (2013) Computational methods and clinical applications for spine imaging. In: Jianhua Y, Tobias K, Shuo L (eds) Vertebrae detection and labelling in lumbar MR images. Springer, Cham, pp 219–230

29. Yang D, Xiong T, Xu D, Huang Q, Liu D, Zhou SK, Xu Z, Park J, Chen M, Tran TD, Chin SP, Metaxas D, Comaniciu D (2017) Automatic vertebra labeling in large-scale 3D CT using deep image-to-image network with message passing and sparsity regularization. In: Information processing in medical imaging (IPMI), pp 633–644

30. Glocker B, Zikic D, Konukoglu E, Haynor DR, Criminisi A (2013) Vertebrae localization in pathological spine CT via dense classification from sparse annotations. Med Image Comput Comput Assist Interv MICCAI 2:262–270

31. Chen H, Shen C, Qin J, Ni D, Shi L, Cheng JCY, Heng P-A (2015) Automatic localization and identification of vertebrae in spine CT via a joint learning model with deep neural networks. Med Image Comput Comput Assist Interv MICCAI 1:515–522