
Article:
Yao, J., Burns, J. E., Forsberg, D. et al. (13 more authors) (2016) A multi-center milestone study of clinical vertebral CT segmentation. Computerized Medical Imaging and Graphics, 49. pp. 16-28. ISSN 0895-6111

https://doi.org/10.1016/j.compmedimag.2015.12.006

A Multi-center Milestone Study of Clinical Vertebral CT Segmentation

Jianhua Yao[1], Joseph E. Burns[2], Daniel Forsberg[3], Alexander Seitel[4], Abtin Rasoulian[4], Purang Abolmaesumi [4], Kerstin Hammernik[5], Martin Urschler[6], Bulat Ibragimov[7], Robert Korez[7], Tomaž Vrtovec[7], Isaac Castro-Mateos[8], Jose M. Pozo[8], Alejandro F. Frangi[8], Ronald M. Summers[1], Shuo Li[9*]

1. Imaging Biomarkers and Computer-Aided Detection Laboratory, Radiology and Imaging Sciences, National Institutes of Health Clinical Center, Bethesda, MD 20892, USA
2. Department of Radiological Sciences, University of California, Irvine, CA 92868, USA
3. Sectra, Linköping, Sweden & Center for Medical Image Science and Visualization (CMIV), Linköping University, Sweden
4. Department of Electrical and Computer Engineering, University of British Columbia, Vancouver, BC, Canada
5. Institute for Computer Graphics and Vision, BioTechMed, Graz University of Technology, Austria
6. Ludwig Boltzmann Institute for Clinical Forensic Imaging, Graz, Austria
7. University of Ljubljana, Faculty of Electrical Engineering, Ljubljana, Slovenia
8. Centre for Computational Imaging and Simulation Technologies in Biomedicine (CISTIB), Department of Mechanical Engineering, University of Sheffield, UK
9. GE Healthcare & University of Western Ontario, Canada

Abstract: A multiple center milestone study of clinical vertebra segmentation is presented in this paper. Vertebra segmentation is a fundamental step for spinal image analysis and intervention. The first half of the study was conducted in the spine segmentation challenge in 2014 International Conference on Medical Image Computing and Computer Assisted Intervention (MICCAI) Workshop on Computational Spine Imaging (CSI 2014). The objective was to evaluate the performance of several state-of-the-art vertebra segmentation algorithms on computed tomography (CT) scans using ten training and five testing dataset, all healthy cases; the second half of the study was conducted after the challenge, where additional 5 abnormal cases are used for testing to evaluate the performance under abnormal cases. Dice coefficients and absolute surface distances were used as evaluation metrics. Segmentation of each vertebra as a single geometric unit, as well as separate segmentation of vertebra substructures, was evaluated. Five teams participated in the comparative study. The top performers in the study achieved Dice coefficient of 0.93 in the upper thoracic, 0.95 in the lower thoracic and 0.96 in the lumbar spine for healthy cases, and 0.88 in the upper thoracic, 0.89 in the lower thoracic and 0.92 in the lumbar spine for osteoporotic and fractured cases. The strengths and weaknesses of each method as well as future suggestion for improvement are discussed. This is the first multi-center comparative study for vertebra segmentation methods, which will provide an up-to-date performance milestone for the fast growing spinal image analysis and intervention.

* Corresponding author. Dr. Shuo Li, Email address: slishuo@gmail.com
The participating authors are ordered by the time their methods were submitted to the comparative study.

1. Background

The vertebral column, also known as spine, is a bony skeletal structure forming the central weight-bearing axis of the human upper body. Multiple medical imaging modalities, such as radiographs, CT, MRI and PET, are used to evaluate spine anatomy and diagnose spinal pathology. Using current generation of scanning techniques, CT is the most spatially accurate modality to assess the three dimensional morphology of the vertebra. Spine segmentation is a fundamental step for most subsequent spine image analysis and modeling tasks, such as identification of spine abnormalities (e.g. vertebral fractures, [1]), image-based biomechanical modeling (e.g. load analysis [2]) or image-guided spine intervention (vertebral fusion, [3]). The accuracy of the segmentation is demanded in some analysis. For instance, image-guided spine intervention often requires sub-millimeter precision. Manually segmenting a vertebra is time consuming and subjective. Fully automated or semi-automated methods are required for most clinical applications.

Vertebra segmentation is challenging due to the complex shape and variable architecture of vertebrae across the population, similar structures in close vicinity, pathology, and the spatial inter-relation between vertebrae and ribs. In recent years, a number of spine segmentation algorithms for computed tomography (CT) images have been proposed. In early work, segmentation of vertebrae was achieved by unsupervised image processing approaches such as adaptive thresholding, region growing and boundary adjustment (Kang et al. [4]), or region-based segmentation such as watershed (Li. et al. [5]) and graph-cut (Aslan et al. [6]). Level set methods had also been adopted since they can handle the complex topological merging and breaking in the vertebrae. Lim et al. [7] included the Willmore flow in a level set framework to guide a surface model evolution. Huang et al. [8] combined edge- and region- based level set functions for vertebra segmentation on CT images. Li et al. [9] proposed an automatically initialized level set method based on hybrid morphological filter and Gaussian mixture model to deal with the topological variation. In region-based techniques, Blumfield et al. [10] devised a statistical and heuristic methods to detect key features for vertebral body segmentation. Yao et al. [11] presented a technique based on watershed algorithm, directed graph search, curved reformation and vertebra template to automatically partition and segment the spinal column. Naegel et al. [12] applied mathematical morphology and watershed for the labeling and segmentation of vertebrae.

More recent methods were mostly based on geometric models, statistical anatomical models, or probabilistic atlas. The models incorporated prior knowledge about the vertebra anatomy. Furthermore, the statistical models estimate the mean shape and variation of a vertebra from a training set of segmented vertebrae. The models were fit to the target image data either through forces derived from the image or via a deformable registration framework. These models are often sensitive to the initial pose estimation, which are done either manually or automatically. The manual initialization may be performed by placing seeds within the vertebral body [13] or drawing bounding box to restrict the searching range [14]. Automatic initialization had also proposed via detecting the spine curvature and inter-vertebral disk [11]. Klinder et al. [15] proposed a method, by integrating detection, identification, and segmentation of vertebrae in a single framework. The method was based on spinal curve extraction and statistical shape models (SSM). The method proposed in Ma et al. [16] was based upon bone-structure edge detectors and coarse-to-fine registration of a deformable surface model for the thoracic spine. Both shape and pose statistics were incorporated in Rasoulian et al. [17] in a multi-vertebrae model for lumbar spine segmentation. Kim and Kim [18] proposed a deformable fence model to separate lumbar vertebrae and

surrounding tissues. Individual vertebrae was modeled in an articulated spine model with a low-dimensional manifold representation and inferred the model using high-order Markov random fields (Kadoury et al. [19]) . The vertebrae were clustered into sub-groups using manifold learning and a linear point distribution model was constructed for each sub-group. Ibragimov et al. [20] built landmark-based shape representations of vertebrae using transportation theory and aligned the model to a specific vertebra in three-dimensional (3D) CT images using game theory. Part-based models and active shape models were used in Roberts et al. [21] to divide the vertebra into several parts and conducted the segmentation collaboratively. The model was applied on 2D radiograph images and can be extended to 3D. Stern et al. [22] proposed a 3D superquadric model for the segmentation of just the vertebral body. Multi-atlas with joint label-fusion had showed promising results in the segmentation of several anatomical organs including vertebrae. Wang et al. [23] applied the atlas approach in the segmentation of osteoporotic vertebrae with compression fractures. Ghebreab and Smeulders [24] constructed a deformable integral spine model encoded as an necklace model by learning the appearance of vertebrae boundaries from a set of training images.

More recently, machine learning techniques had been applied in the segmentation of vertebrae. Huang et al.[25] applied a statistical learning approach based on Adaboost for vertebra detection and an iterative normalized cut algorithm for boundary refinement. Suzani et al. [26] proposed a deep learning scheme to automatically localize, identify and segment vertebral body in MR images. Mirzaalian et al. [27] combined a probabilistic boosting tree classifier for initialization of statistical shape models for segmentation.

Most of the published methods reported fairly accurate results (1.12 ± 1.04 mm point-to-surface error reported in Klinder et al. [15]). Table 1 summarizes the performance of some recently published methods, including number of cases, performance metrics by DICE coefficient and average surface distance, targeted subjects, and initialization methods. All information is directly extracted from the published peer-reviewed papers. However, these algorithms were mostly evaluated on different data sets with various degrees of difficulties and are not publicly available. This precludes direct comparison of the results and access to the data, and therefore, their performances were not independently verified.

In order to objectively compare different segmentation algorithms, it is necessary to establish standardized reference data and validation criteria. Quite a few challenge frameworks have been developed in the past few years for several medical image analysis problems. A dedicated website at http://www.grand-challenge.org has the most complete list of organized challenges since 2007, including those focusing on liver, lung, brain and heart, amongst others. Thus far, there is no grand challenge focusing specifically on spine image analysis.

This paper presents a milestone comparative study of the vertebra segmentation. The first stage was a challenge held at the 2014 International Conference on Medical Image Computing and Computer Assisted Intervention (MICCAI) Workshop on Computational Spine Imaging (CSI 2014), and the second stage was evaluation on more challenging clinical cases. The objective to organize this vertebra segmentation comparative study was three-fold. Firstly, we wanted to provide a platform to objectively evaluate the strengths and weaknesses of various spine segmentation algorithms; secondly, we intended to construct an annotated reference data set for spine labeling and segmentation; and thirdly, we want to assess current state-of-the-art segmentation accuracy for vertebra and its substructures. The details of the organization of the comparative study can be found at the website (http://csi-workshop.weebly.com/).

Table 1. Summary of recently published vertebra segmentation methods on CT

Method	Number of Cases	DICE	Average Surface Distance (mm)	Subject	Initialization
Kang (2003)	3	N/A	18.36	Lumbar phantom	Manual
Klinder (2009)	64	N/A	1.12	Whole spine	Auto
Kim (2009)	50	87%	N/A	Lumbar	Auto
Huang (2009)	22	96%	N/A	Vertebral body	Auto
Aslan (2010)	30	94.4%	N/A	Lumbar Vertebral body	Auto
Stern (2011)	150	N/A	1.17	Lumbar Vertebral body	Manual
Kadoury (2011)	711	N/A	1.8	Whole spine	Auto
Ibraginov (2011)	50	93.6%	0.75	Lumbar	Auto
Huang (2013)	56	94%	N/A	Lumbar	Manual
Ma (2013)	40	N/A	0.95	Thoracic	Auto
Mirzaalian (2013)	154	N/A	1.37	Whole spine	Manual
Rasoulian (2013)	32	N/A	1.38	Lumbar	Manual
Lim (2013)	20	89.3%	N/A	Lumbar	Manual
Suzani (2015)	9	N/A	2.8	Lumbar Vertebral body	Auto
Li (2015)	25	91.7%	7.73	Lumbar	Auto
Wang (2015)	170	92.7%	0.32	Whole spine	Auto

2. Spine imaging data sets

The data sets used in the comparative study were acquired at the University of California, Irvine, Medical Center (Orange, CA, USA), between March 2013 and May 2013. The study received Institutional Review Board approval, and was compliant with the Health Insurance Portability and Accountability Act. As the

study was performed as a retrospective analysis of previously obtained imaging studies, informed consent was waived. The data sets were manually selected by a radiologist with eight years of experience according to the following selection criteria: thoracic and lumbar spine column scanned. All patients were scanned using a spine CT protocol, where a small field of view centered at the spine was reconstructed. The scanning parameters included 0.7-2.0 mm slice thickness, 120 kVp, soft tissue reconstruction kernel, and intravenous contrast. The volumes completely covered the thoracic and lumbar part of the spine and were scanned as a single continuous CT data set at high spatial resolution.

Following these data characteristics, we collected twenty data sets for the comparative study, ten for training and ten for testing. The training cases were provided before the participants entered the workshop challenge (January, 2014). The testing sets were provided in two stages after the participants entered the challenge. In the first stage, five cases from healthy young individuals (20-34 years, mean 27 years) were provided. In the second stage, the participants were invited back to test on five cases from an osteoporotic cohort (59-82 years, mean 73 years) that has been previously identified to have at least one vertebral compression fracture. In the osteoporotic set, 16 vertebrae were identified with a compression fracture (one with grade 1 Genant score, ten with grade 2, and five with grade 3). Examples of the two testing sets are shown in Figure 1. All data were anonymized and made available in Meta format (MHD/raw). The data sets and related codes are open to public and released on SpineWeb (http://spineweb.digitalimaginggroup.ca/spineweb/index.php?n=Main.Datasets), a collaborative platform for research on spine imaging and image analysis. The details of the training and test data sets are listed in Table 2. The test sets from healthy young individual were intended to assess the baseline performance and the more difficult diseased cases for evaluating the accuracy of state-of-the-art vertebra segmentation algorithms.

Table 2. Description of training and test data sets

Case	Gender	Age	Manufacturer	Model	Pixel spacing(mm)	Slice thickness (mm)	#slices
Training cases							
case1	F	23	Philips	iCT 256	0.3125	1.0	559
case2	F	22	Philips	iCT 256	0.3125	1.0	507
case3	M	27	Philips	iCT 256	0.3125	1.0	560
case4	M	28	Philips	iCT 256	0.3535	1.0	625
case5	F	19	Philips	iCT 256	0.3125	1.0	601
case6	M	26	Philips	iCT 256	0.3437	1.0	562
case7	F	21	Philips	iCT 256	0.3125	1.0	509
case8	F	16	Philips	iCT 256	0.3125	1.0	548
case9	F	23	Philips	iCT 256	0.3613	1.0	572
case10	F	25	Philips	iCT 256	0.3125	1.0	552
Testing cases							
case1	F	25	Philips	iCT 256	0.3125	1.0	545
case2	M	32	Philips	iCT 256	0.3457	1.0	618
case3	F	34	Siemens	Sensation 64	0.3125	0.7	766

case4	M	25	Philips	iCT 256	0.3125	1.0	551
case5	M	20	Siemens	Sensation 64	0.3125	0.7	938
case6	M	82	Philips	iCT 256	0.3496	1.0	536
case7	F	59	Philips	iCT 256	0.3184	2.0	278
case8	F	68	Philips	iCT 256	0.3516	2.0	265
case9	M	79	Philips	iCT 256	0.3516	2.0	268
case10	F	78	Philips	iCT 256	0.3789	2.0	237

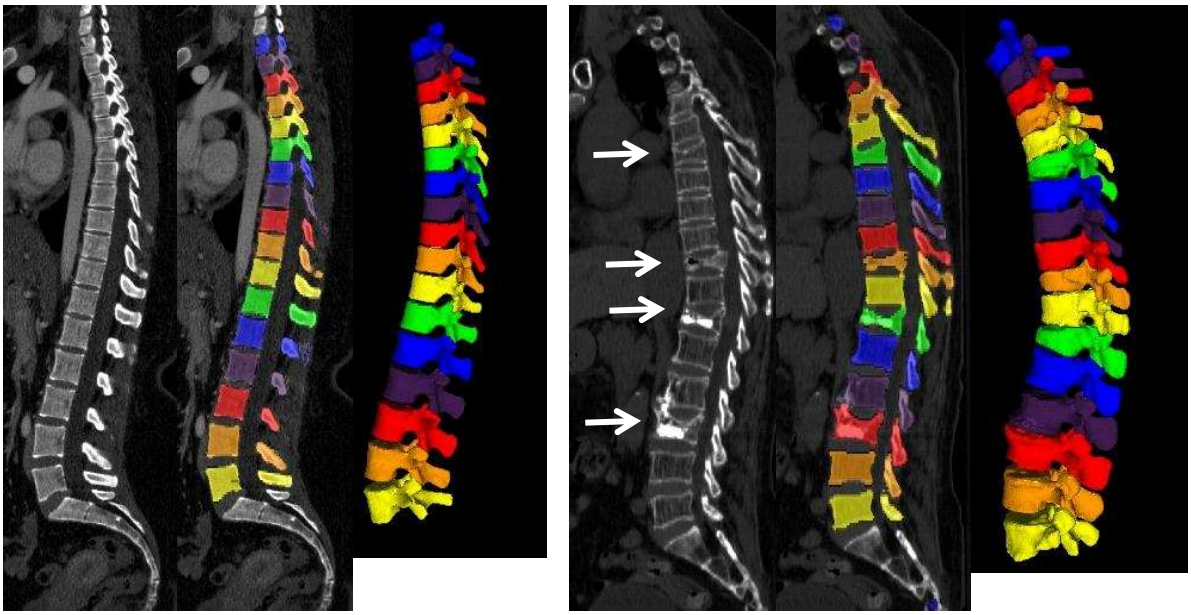


Figure 1. Example of test cases with reference segmentation
Each vertebra is assigned a unique label (color coded). Sagittal and 3D views are shown.
Left: a healthy spine (test case 1)
Right: an osteoporotic and fractured spine (test case 10). Arrows point to vertebrae with compression fractures.

3. Reference data generation

The reference segmentation data was generated in two stages. Firstly, the initial segmentations were obtained using a fully automatic algorithm reported in [11], which was based on adaptive thresholding, watershed, directed graph search, and connected component analysis. The obtained segmentations were then manually corrected and refined by a medical fellow and a research fellow using customized software that was developed for the manual correction. In the reference data, each vertebra was assigned a unique label and the background (pixels other than vertebrae) was assigned label 0. We created reference segmentation for every vertebra, for both thoracic (T1-T12) and lumbar (L1-L5) spines. A reference segmentation file was saved in Meta data format [28] for each data set with the same resolution as the original CT image file. The reference segmentations for the test sets were based on consensus reading of

two operators. Figure 1 shows examples of the reference labels in the sagittal plane and the 3D surface model generated by the reference segmentation for both a healthy case and an osteoporotic case.

4. Participating algorithms

Since the release of the training data there have been over 60 requests or downloads of the training data from SpineWeb. Five teams entered the comparative study held at the Computational Spine Imaging Workshop (CSI2014). Among the participants, four teams segmented both thoracic and lumbar vertebrae, and one team segmented only lumbar vertebrae.

The five participating algorithms are dubbed as Method 1 [29], Method 2 [30], Method 3 [31], Method 4 [32], and Method 5 [33] in this paper according to the order of the submission. The titles of the five methods are listed in the references. The following is a brief description of each participating method.

Method 1 [29] is an atlas-based technique and consists of four steps: pre-processing, initial alignment, non-rigid registration and label fusion. N spine atlases (image data with corresponding label data) are used to segment thoracic and lumbar vertebrae as imaged in a target data set. In the pre-processing step, consisting of spinal canal tracking, disc detection, and vertebra position and rotation estimation, an approximate position and rotation (pose) of each vertebra in all data sets are estimated. The results from the pre-processing are used to obtain an initial alignment between each of the N atlases and the target data set. The initial alignment can either be for the whole spine or computed per vertebra. This is followed by a registration step, where the vertebrae of each atlas are registered to the target data set using non-rigid registration (minimizing the local phase-difference). This step is performed on a group of three vertebrae. The computed transforms are used to transform corresponding label sets. The transformed labels are combined to a single label volume using label fusion (majority voting) to form the segmentation of the vertebrae as imaged in the target data set.

Method 2 [30] is based on a statistical multi-vertebrae shape+pose model which is registered to the bony edges of the spinal column as extracted from the CT volume. For construction of the model the idea is to analyze the pose and shape statistics separately as they are not necessarily correlated and are not formulated in the same parameter space. Training data is collected for every vertebra (in this case T1 to L5) and is used to build individual sub-models each containing three neighboring vertebrae and the ensemble of all models covering the whole spinal column. Segmentation using one of these statistical multi-object shape+pose model can then be formulated as a registration problem where the model is registered to the bone edge point cloud extracted from the CT volume by optimizing the model parameters using the expectation maximization (EM) algorithm. Successive registration of the individual sub-models starting from a user-specified initial vertebra finally results in the segmentation of the spine.

Method 3 [31] uses a variational segmentation framework which is derived from a convex formulation of 3D geodesic active contours for individual vertebra segmentation. In this formulation the weighted total variation (TV) norm is combined with prior bone intensity and shape information. For bone intensity prediction, normalized foreground (bone) and background histograms are learned from annotated training data. Each voxel of a test image is classified accordingly using the log-likelihood ratio. Prior shape knowledge in terms of a mean shape is obtained by registering the set of binary training vertebrae representations and final averaging. To account for variation in shape along the spine, mean shape representations for upper thoracic, lower thoracic and lumbar spines are calculated separately. At testing time, the learned mean shape is registered to a binary representation of the bone prior to get a rough location of the vertebrae. The final energy formulation expressed in a variational framework

obtains a segmentation by combining bone prior, registered shape prior and the weighted TV norm which accounts for both edge magnitude and edge direction of the respective image.

Method 4 [32] consists of two parts: vertebra detection and vertebra segmentation. An interpolation-based optimization approach is applied to detect the whole spine and individual vertebrae in an unknown CT spine image by using a spline-based interpolation function on an equidistant sparse optimization grid and a dimension-wise computational complexity reduction algorithm to obtain the optimal translation, scaling and rotation parameters of the rigidly align vertebra shape models. The obtained detection results represent a robust and accurate initialization for vertebra segmentation, built upon the existing shape-constrained deformable model approach. The proposed iterative segmentation consists of finding distinctive vertebra boundaries by applying Canny edge operator and random forest regression model of image intensities and intensity gradients, and of deforming the vertebra shape model so that it fits the obtained vertebra boundaries while preserving shape topology.

In Method 5 [33], a statistical shape models (SSM) of each lumbar vertebra was previously created from an independent dataset of 30 lumbar spines with no evident osteological pathologies. From manually placed intervertebral discs centers, the similarity transformation parameters of each vertebra are computed to initialize the vertebra shapes. The segmentation is performed by iteratively deforming a mesh inside the image intensity and then projecting it into the SSM space until convergence. Afterwards, a relaxation step based on B-spline is applied to overcome the SSM rigidity. The deformation of the mesh, within the image intensity, is performed by displacing each landmark along the normal direction of the surface mesh at the landmark position seeking a minimum of a cost function based on a set of trained features.

The technical comparisons of the five methods are detailed in Table 3. The comparisons are conducted in six aspects: vertebral localization, segmentation strategy, bundled model, registration/optimization, image feature, and running time. Method 1 and Method 4 comprise steps to automatically initialize the location of vertebrae. Method 2, 3, and 5 require manual initialization of the model, either at the center of vertebral bodies or at the center of intervertebral discs. The automatic initializations of Method 1 and 4 were self-claimed by the participants and not verified. All methods are based on certain types of shape and intensity models. Method 1 uses multiple atlases directly derived from the reference data. Method 2 uses a statistical shape+pose model built from 87 training volumes incorporating variations of both shape and pose across the population. Both Method 3 and Method 4 use mean shape models from the training data. Method 4 builds a model for each vertebra level, while Method 3 builds one model for each section of the spine (one for T1-T6, one for T7-T12, and one for L1-L5). Method 5 computes a statistical shape model from 30 training models for each vertebra level. In Method 1 and Method 2, adjacent vertebrae are bundled together (5 vertebrae in Method 1 and 3 vertebrae in Method 2) in the segmentation. Different registration/optimization frameworks were adopted in the methods. Method 1 first applies a non-rigid registration for each atlas and then performs a label fusion. Method 2 conducts an EM algorithm to optimize the model. Method 3 adopts a total variation framework. Surface mesh deformation and reconfiguration is performed in Method 4. Method 5 employs statistical shape deformation plus B-spline relaxation for surface optimization. Different feature functions are used in the methods for the optimization. They are mostly based on edge point and intensity models. The running time was reported by the participating teams (Table 3). It can only be viewed as a reference, since the algorithms were run on different hardware platforms and some have been optimized by graphics processing units (GPU).

Table 3. Detailed comparison of five participating algorithms.

Method	Vertebral Localization	Segmentation Strategy	Bundled model	Registration/Optimization	Image Feature /Energy Formula	Runtime
1 [29]	Automatic	Multi-atlas	Five vertebrae	Nonrigid+label fusion	Local phase difference	12 minutes per case (GPU)
2 [30]	Manually specify the vertebral body center	Statistical multi-object shape+pose model (87 training volumes)	Three vertebrae	EM algorithm	Edge point data / distance of edge points	10 minutes per case
3 [31]	Manually specify the vertebral body center	Mean shape model (three model groups)	N/A	Total variation framework	Bone prior map and image edge influence/ total variation of edge magnitude and direction	45 minutes per case (GPU)
4 [32]	Automatic (interpolation theory)	Mean shape model (each vertebra)	N/A	Mesh deformation	Canny edge +random forest intensity regression/ boundary difference	30 minutes per case
5 [33]	Manually specify the center of the intervertebral disc below the vertebra	Statistical shape model (each vertebra) (30 training models)	N/A	Statistical shape deformation + B-spline relaxation	Image intensity, directional derivative/ surface feature difference	10 seconds per vertebra

5. Evaluation

The performance on the training data set was evaluated by the participants themselves and reported in their submissions [29-33]. The performance on the test data set was evaluated by the organizers.

After the test data set was released, the participants were given 10 days to submit the segmentation results. Each segmented vertebra was assigned a unique label. The results were submitted in Meta format (MHD/raw) [28] with the same resolution as the original CT data.

Two metrics were employed for evaluation: Dice coefficient (DC) [34] and mean absolute surface distance (ASD). The definitions are as follows:

$$DC = \frac{2|\mathrm{Vr} \cap \mathrm{Vs}|}{|\mathrm{Vr}| + |\mathrm{Vs}|}$$

$$ASD = \frac{1}{|S_s|}\sum_{i=1}^{|S_s|} \|d_i(S_s, S_r)\|$$

Here V_r is the reference volume, V_s is the segmentation volume, S_r is the reference surface, S_s is the segmentation surface, and d_i is the minimum distance from a point on S_s to S_r. The evaluation was conducted on each individual vertebra. The maximum surface distance (MSD) is also evaluated. The surfaces were generated using a Marching Cube algorithm from the binary segmentation mask.

By visual inspection, we noticed that the segmentation performed differently at different parts of the vertebra. Therefore, we evaluated the segmentation performance for both the whole vertebra and its four substructures: vertebral body, left transverse process, right transverse process, and spinous process, respectively. We developed an automatic method to partition the vertebra into anatomical substructures [35]. It is based on the anatomical knowledge that pedicles and laminae are the densest parts of the vertebral arch which forms the circle of bones around the spinal canal. We therefore search for the four cutting planes at left pedicle, right pedicle, left lamina and right lamina that go through the cross-sections with highest CT intensity around the spinal canal. Symmetric constraints are also enforced to balance the left and right cutting planes. The vertebra is then partitioned into four substructures (vertebral body, left transverse process, right transverse process and spinous process), depending on which side of the cutting planes a pixel lies. The partitions on our data sets were verified by experts to ensure the correctness. Figure 2 shows the partitioning of a vertebra into the four substructures.

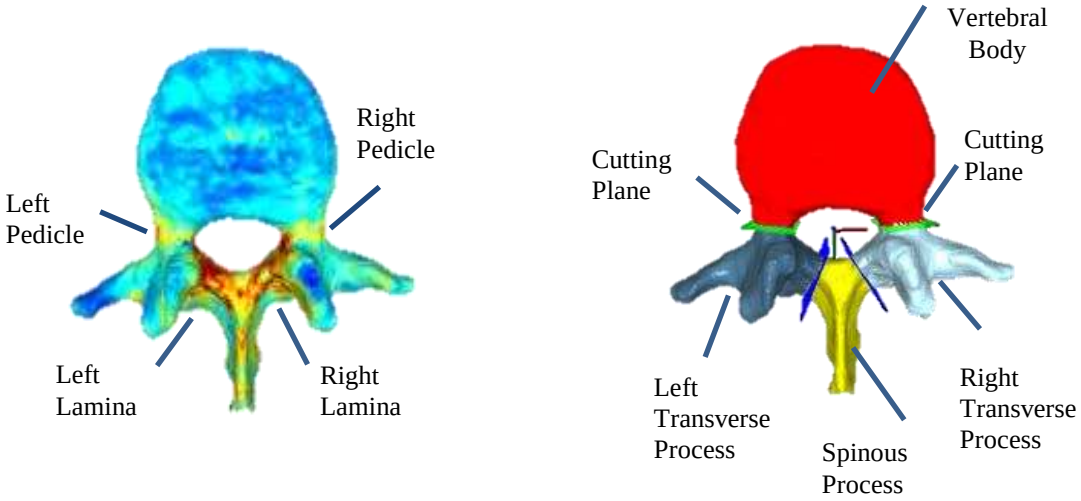


Figure 2. Partitioning of a vertebra into four substructures

Left: Density map on vertebra surface, hotter color: higher density
Right: Partitioning a vertebra into four substructures. The substructures are color-coded with different colors. The cutting planes lie at the border between two substructures.

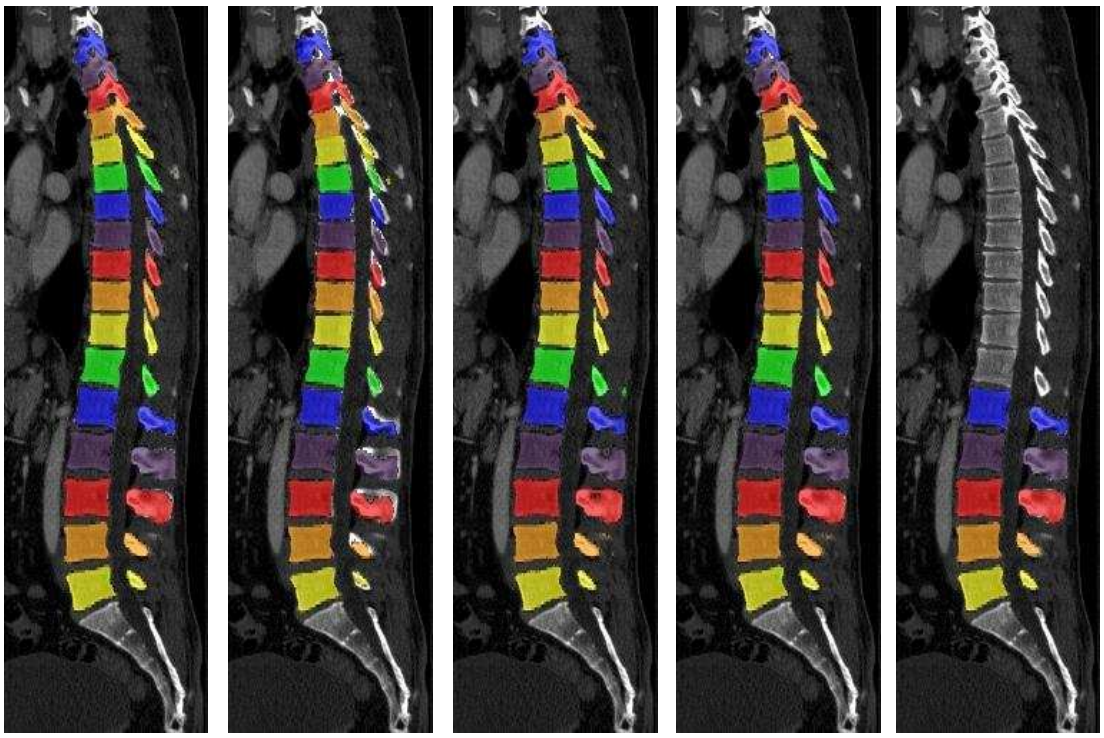
6. Performance comparison

The segmentation results were compared both visually and quantitatively. The results were superimposed on the CT image for visual inspection. DC and ASD were used for quantitative analysis. In this paper, we mainly focus on the results on the test set.

Figure 3 shows the visual comparison of submitted segmentation results for test case 1 from a healthy spine. All methods achieve visually acceptable segmentation for thoracic and lumbar vertebrae in a healthy spine. There is no obvious leakage or under-segmentation from the sagittal view. Figure 4 shows the visual results for test case 10 which is an osteoporotic case with multiple compression fractures. All methods show certain degree of deterioration in performance compared to the healthy case. Method 1 and 4 demonstrated the best segmentation on the osteoporotic case. Method 1's segmentation on the

compression fractured vertebrae slightly leaked into the intervertebral disc space. Method 4 slightly under-segmented the spinous processes. Method 2 had trouble locating the endplates in the thoracic spine. Method 3 failed to segment several vertebrae because of the training prior. Method 5 only segmented the lumbar spine, but showed fairly good results, not statistically different from method 4 for osteoporotic spines.

For a closer visual inspection, Figure 5 shows the visual comparison of the segmentation of the mid-axial slice for three representative vertebrae on a healthy case: T3, T9 and L3. In T3 and T9, all methods successfully separate the vertebra and the ribs. The border of segmented vertebra in Method 1 is not smooth, which indicates that further refinement is necessary. Method 1 participant indicated that the data was resampled at 1mm × 1mm × 1mm due to memory limitation. Running the method on a finer grid may improve the pixelated result. The segmentation in Method 2 is off-mark although the location of the vertebra and the overall shape are correct. Another stage of local segmentation should be conducted. Method 3 and 4 both achieve moderately accurate segmentation results, but it is noted that the segmentation of the posterior substructures still have room for improvement. The tips of the processes are not completely segmented and some contrast-enhanced vessels are included in the segmentation. Method 5 only segments the lumbar spine and the result is similar to that of Method 1 where the boundary is slightly off.



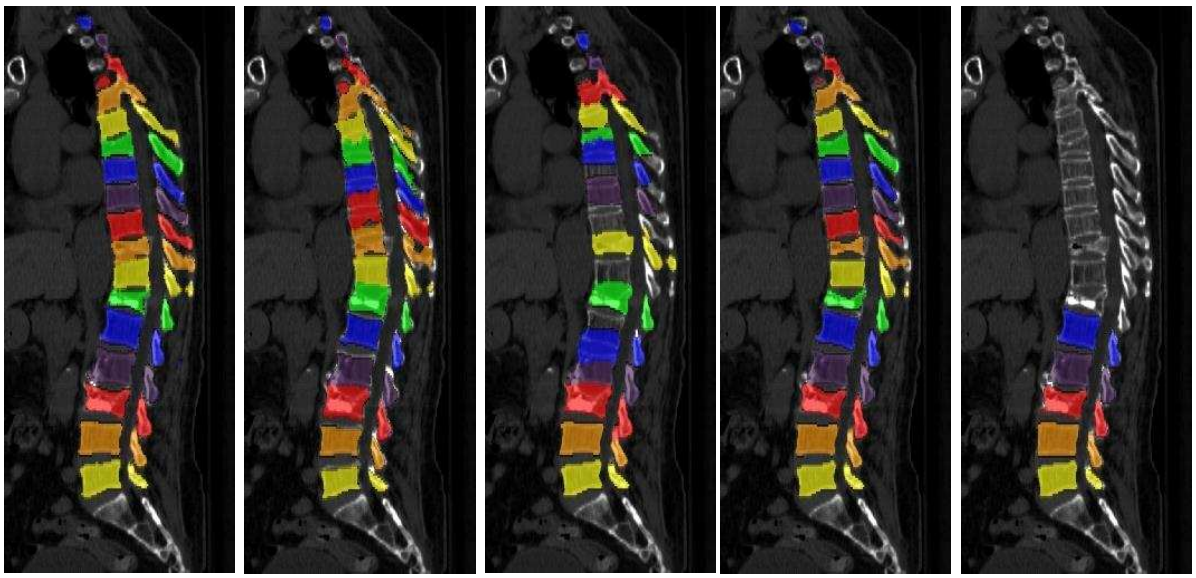
Method 1 Method 2 Method 3 Method 4 Method 5

Figure 3. Visual comparison of segmentation results for test case 2 (a healthy case)

There is a general trend of better performance from upper spine to lower spine as the vertebrae gradually increase in size and density. To illustrate the pattern, we group the vertebrae into three sections: upper thoracic from T1 to T6, lower thoracic from T7 to T12 and lumbar spine from L1 to L5 (Figure 6).

In the healthy cases, DC goes from 0.867 in the upper thoracic, to 0.909 in the lower thoracic and to 0.933 in the lumbar spine. In the osteoporotic cases, DC goes from 0.652 in the upper thoracic, to 0.756 in the lower thoracic and to 0.854 in the lumbar spine. As expected, the performance on the healthy cases is much higher than that on the disease cases.

Figure 7 summarizes the performance on the sub-structures. This evaluation was only conducted on the healthy cases. It is noted that the DC for the vertebral body segmentation is much higher than that for the posterior substructures (left transverse process, right transverse process and spinous process). The three processes have comparable performance. This further verifies the visual comparison shown in Figure 5.



Method 1 Method 2 Method 3 Method 4 Method 5

Figure 4. Visual comparison of segmentation results for test case 10 (an osteoporotic and fractured case).

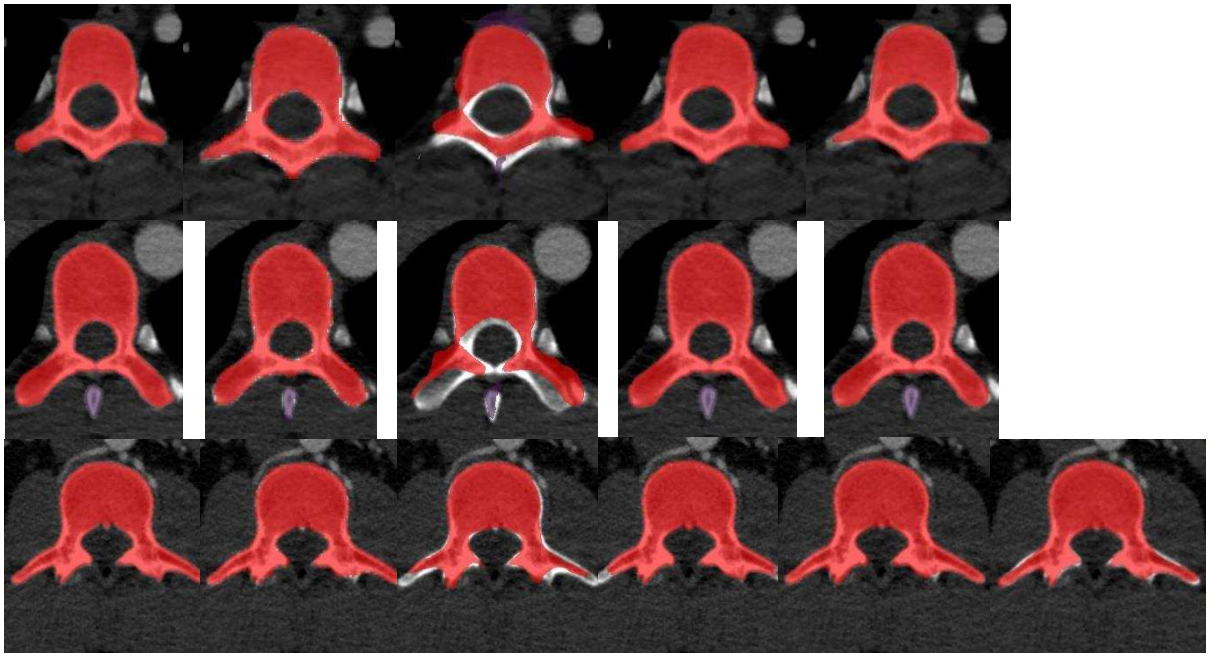


Figure 5. Visual comparison of segmentation results for specific vertebrae in test case 4
Row 1: T3 vertebra; Row 2: T9 vertebra; Row 3: L3 vertebra
Mid-axial slice for each vertebra is shown. The segmentation is superimposed on the CT data.

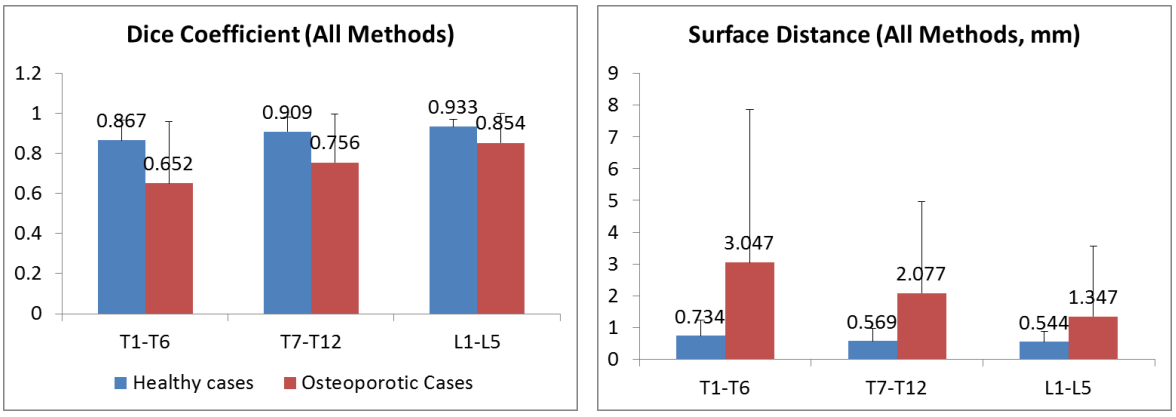


Figure 6. Mean performance of all methods for healthy and osteoporotic cases.
Both Dice Coefficient and mean surface distance are evaluated. The spine column is divided into three segments (T1-T6, T7-T12, and L1-L5) for assessment.

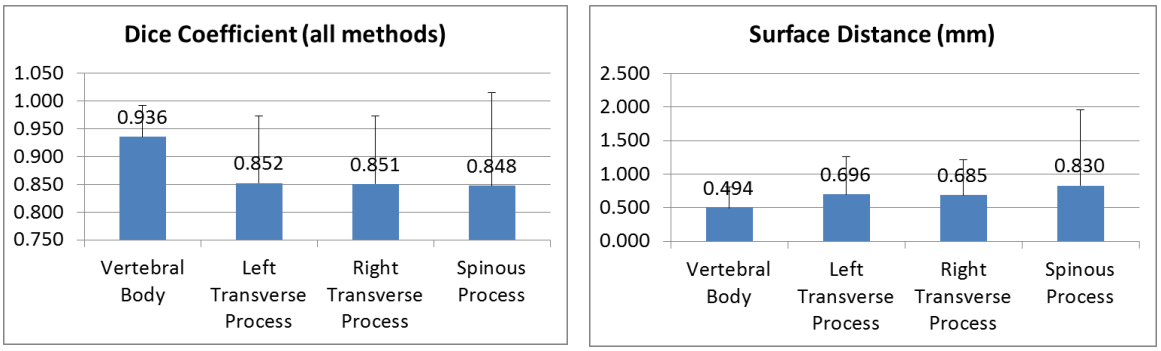


Figure 7. Mean performance of all methods for vertebra substructures.
The results show that vertebral body is more accurately segmented than other posterior sub structures.

Figure 8 shows the comparison for the three vertebra groups on both the healthy and osteoporotic cases for all five methods. Figure 9 summarizes the comparison for the whole thoracic and lumbar spine. These comparisons show the differences in performance among the methods, in terms of both DC and ASD. The MSD for healthy cases were 8, 12, 22, 12 and 10mm for Method 1 to 5 respectively, and those for osteoporotic cases were 9, 15, 44, 12 and 10mm for Method 1 to 5 respectively. Since DC and ASD show similar patterns, we will only show DC in most of the following comparisons. The ranking of performance on the healthy cases is Method 4, Method 3, Method 1, Method 5, and Method 2, and that on the osteoporotic cases is Method 1, Method 4, Method 5, Method 2 and Method 3. It is noted that Method 5 only segmented lumbar vertebrae. Method 3 performed well on the healthy cases, but failed on many vertebrae in the osteoporotic cases (34 out of 85 vertebrae were not segmented). For those vertebrae that were successfully segmented in Method 3, the average DC was 0.833. The reason for the failure was that the bone prior map was trained based on healthy vertebrae and did not work on most of the osteoporotic vertebrae. Method 5 was also trained on healthy cases, but the B-spline relaxation step seems to give enough flexibility to allow the model to adapt to fractured lumbar vertebrae.

Statistical analysis was conducted to evaluate the significance of differences in performance. Table 4 lists the p-value of the paired t-test on DC of every vertebra for the healthy cases. Only two t-tests did not show statistically significant difference: the comparison between Method 3 and Method 4, and the one between Method 1 and Method 5. The analysis shows that Method 3 and Method 4 statistically performed better than other methods on the healthy cases. Table 5 lists the paired t-test results for the osteoporotic cases. Again only two t-tests did not show statistically significant difference: the comparison between Method 1 and Method 4, and the one between Method 4 and Method 5 (only on lumbar vertebrae). This shows that Method 1 and Method 4 had statistically better performance than other methods on the osteoporotic cases. Table 6 lists the z-test results comparing DC on the healthy cases and the osteoporotic cases for each method. It shows Method 1 had the smallest differences between the two test sets, while all methods showed statistically worst performance on the osteoporotic cases comparing to the healthy cases.

In the osteoporotic cases, 16 out of 85 vertebrae were previously identified with compression fractures. The comparison of performance on fractured and non-fractured vertebrae is shown in Figure 10. All methods except Method 2 had better performance on non-fractured vertebrae than fractured vertebrae.

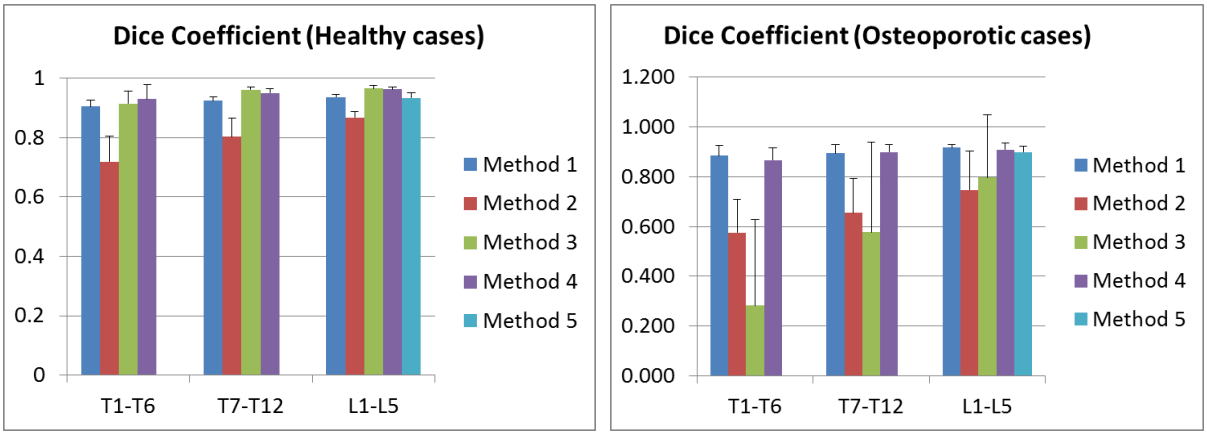


Figure 8. Performance comparison on each vertebra group on healthy (left) and osteoporotic (right) cases. The charts compare the performance of different methods on each segment of the spine.

Table 4. Paired t-test among methods (p-value) on the healthy cases

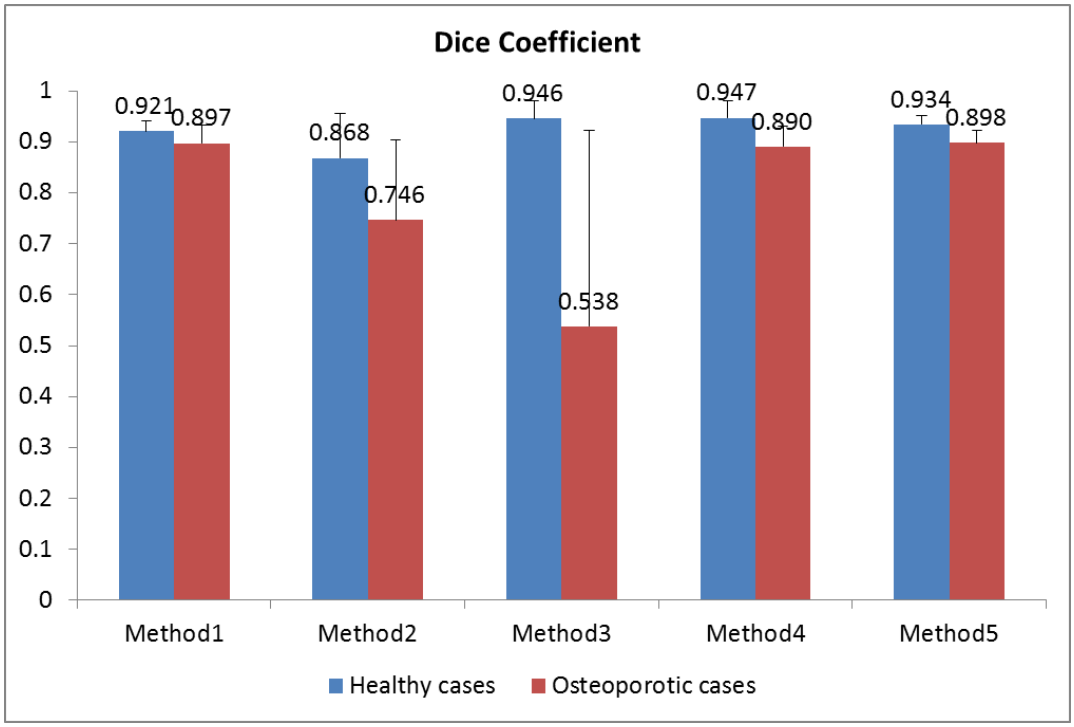
	Method1	Method2	Method3	Method4	Method5
Method1		$<10^{-3}$	$<10^{-3}$	$<10^{-3}$	0.519
Method2			$<10^{-3}$	$<10^{-3}$	$<10^{-3}$
Method3				0.846	0.001
Method4					0.001

Table 5. Paired t-test among methods (p-value) on the osteoporotic cases

	Method1	Method2	Method3	Method4	Method5
Method1		$<10^{-3}$	$<10^{-3}$	0.12	0.004
Method2			0.01	$<10^{-3}$	0.01
Method3				$<10^{-3}$	0.12
Method4					0.25

Table 6. z-test between the healthy cases and osteoporotic cases for each method

	Method1	Method2	Method3	Method4	Method5
Difference	0.024	0.138	0.408	0.057	0.036
z value	2.2	12.8	43.6	5.2	1.8
p-value	0.027	0	0	$<10^{-3}$	0.07



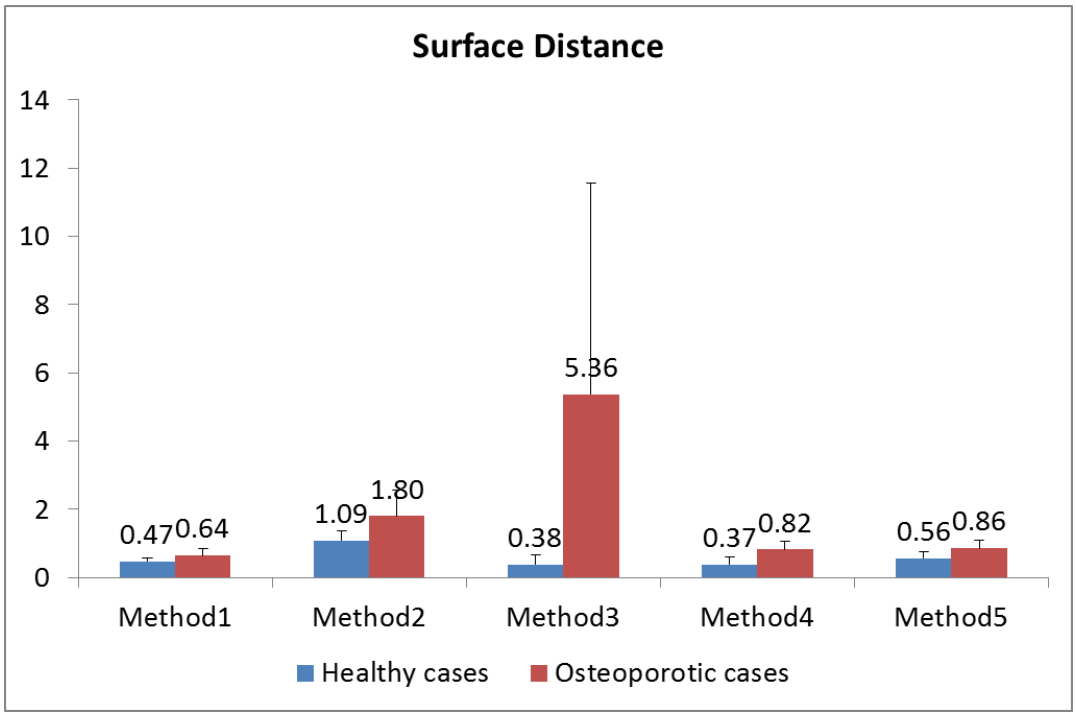


Figure 9. Performance comparison on the entire spinal column

Only lumbar vertebrae were evaluated in Method 5. The charts compare the difference of each method on healthy cases and osteoporotic cases.

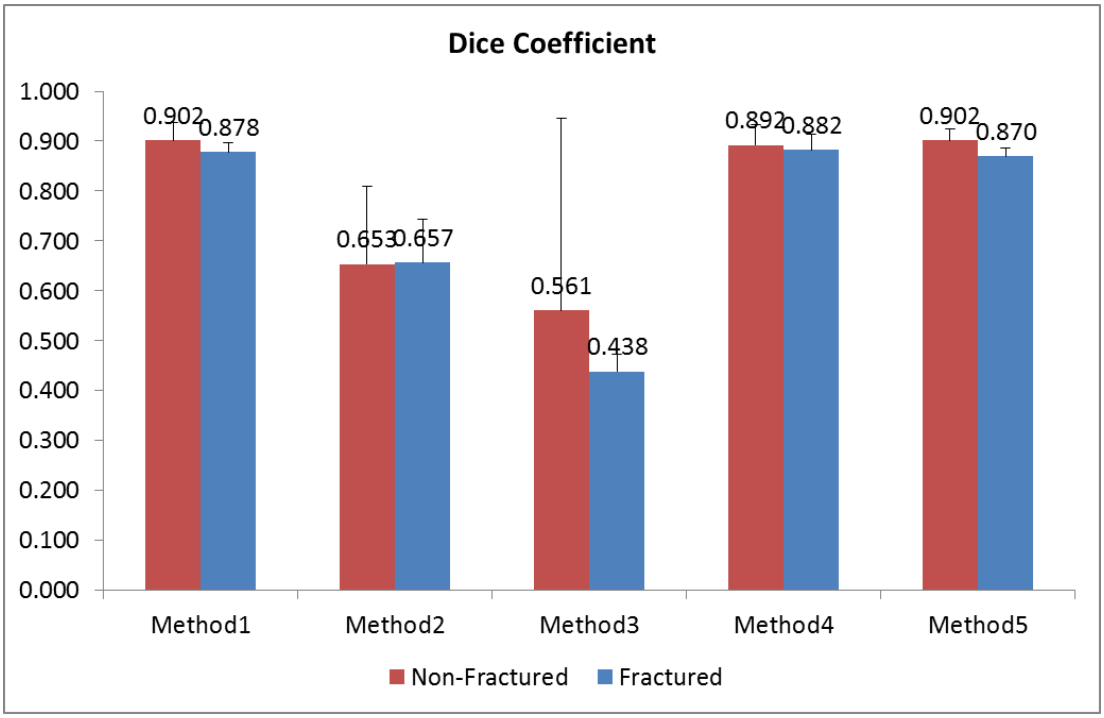


Figure 10. Performance comparison on fractured and non-fractured vertebrae.
All methods except Method 2 performed better on non-fractured vertebrae.

Figure 11 compares the system performance on vertebral substructures for each of the five methods. This evaluation was conducted on the healthy cases. A similar trend in segmentation performance between the different methodologies is seen in the substructure segmentation portion of the comparative study as was seen in the whole vertebra segmentation. All methods perform better on the vertebral body than the other substructures.

We evaluated the inter-operator and intra-operator variability of the manual segmentation to assess the consistency and variability of the reference segmentation. We chose two test data sets (test case 1 for healthy case and test case 6 for osteoporotic case). We then had a second operator to provide a new manual segmentation (for inter-operator variability) and also asked the first operator to repeat the manual segmentation six months after the first manual segmentation (for intra-operator variability). Table 7 lists the mean DC for both data sets between the two corresponding manual segmentations. The manual segmentations showed high consistency even for osteoporotic and fractured vertebrae.

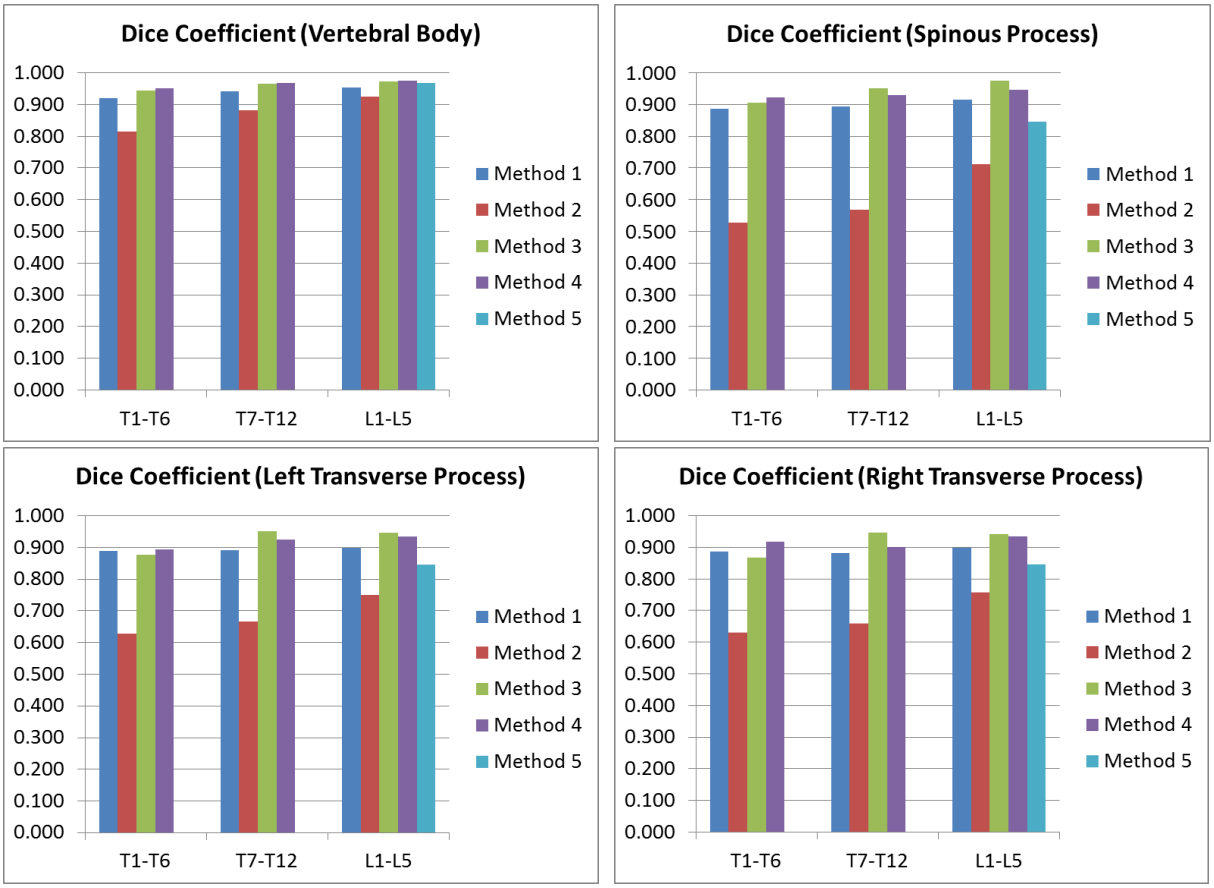


Figure 11. Performance comparison on substructures.
The charts show the similar pattern of performance on substructures.

Table 7. Inter- and intra- operator manual segmentation variability

	T1-T6	T7-T12	L1-L5	All
Inter- operator	0.974±0.005	0.971±0.014	0.983±0.005	0.976±0.01
Intra- operator	0.981±0.004	0.989±0.004	0.993±0.002	0.987±0.006

7. Discussion

The performance of the methods of the participating teams ranged from 0.868 to 0.947 in DC and 0.373 to 1.086 mm in ASD for the healthy cases. The best results represent the state-of-the-art performance and out-perform most recently published methods (Table 1). However, it also indicates that there are still opportunities for improvement. The substructure assessment shows that the vertebral body segmentation is generally excellent (DC of 0.936 on average, 0.964 in the best method). However, for other substructures (left and right transverse processes, spinous process), the DC is 0.852 on average and 0.917 in the best performer. A closer look to Figure 5 also shows that most segmentation errors occur at

the tip of the processes. The tips of the processes are often used as landmarks for image-guided intervention. Further refinement in the posterior substructures is possible and desirable. For the osteoporotic cases, the performance varied greatly among methods (DC from 0.472 to 0.897). Method 1 and Method 4 managed to segment every vertebra, while Method 3 failed for about 40% of the vertebrae.

The results also show that the performance of the segmentation algorithms varies at different vertebra levels. For instance in the healthy cases, upper thoracic levels have a DC of 0.867 on average and 0.930 in best, lower thoracic 0.909 on average and 0.961 in best, and lumbar 0.933 on average and 0.965 in best. The performance difference is predominantly based on two factors: 1) the size and bone density at the upper thoracic level is smaller and lower, respectively, than that at the lumbar level; 2) interfaces with surrounding structures are more complex at the upper thoracic level, particularly at the costovertebral junctions connecting the ribs and the vertebrae. Further investigation is necessary to improve the segmentation of the upper spine column.

All participating methods used models computed from training data to segment the test data. Method 2 and Method 5 used their own training set to build the model. Indeed, the availability of only lumbar training set was the reason for Method 5 to segment only lumbar vertebrae. The difference in imaging protocols and type of population between their own training set and the challenge data set may have decreased the accuracy of the results. Method 1 used multiple atlases from labels, Method 2 and Method 5 used statistical shape models and Method 3 and Method 4 used mean shape models. The mean shape model with large flexibility for deformation may work better for healthy and normal vertebrae. However, statistical models with stricter domain constraints would be necessary for pathological cases where the target shape is far from the mean shape and therefore severe under- or over- segmentation may occur without shape and domain constraints. Method 1 and Method 4 performed well on both the healthy and diseased cases, indicating both statistical model based approaches and single model based deformable registration approaches are valid for reliable vertebra segmentation.

Initial location of the model is essential for the accuracy of segmentation results. Three methods in this comparative study required manually placement of the model locations. Automatic vertebra labeling and localization will be important for the methods to be applied in a clinical setting or to a large number of data sets. Manually placed seeds also have the issue of operator subjectivity. Methods based on spinal canal tracking [11, 36] or based on random forest models [37, 38] have shown robust and promising results to automatically locate the vertebrae.

Vertebra models are bundled in Method 1 and Method 2, so that the interaction between adjacent vertebrae can be employed to assist the segmentation. Since most vertebrae are well separated, the individual vertebra model is able to reliably segment the individual vertebrae independently, within the limitations of normal architectural variation. In the case of pathologic anatomic deformity, especially for cases with compression fractures, it can be helpful to rely on relatively healthy vertebrae in the neighborhood to assist the segmentation of the damaged vertebra. Thus, the bundled model is expected to be beneficial in situations of pathologic architectural deformity. The bundled model can also help prevent overlapping and collision between adjacent vertebrae. This is one of the reasons that Method 1 performed the best on the osteoporotic and fractured vertebrae.

Vertebrae at different spinal levels have different shapes, sizes, and image intensity. For instance, two vertebrae with a large spatial separation within the spinal column, such as an upper thoracic vertebra and a lower lumbar vertebra, show significant morphologic differences. Therefore, it would be a difficult task to accurately characterize all vertebrae with a single model. Based on this characterization, all participating teams employed methods that built different models for different vertebra levels, or at least

for different vertebral groups (Method 3). Vertebra specific models impose anatomical knowledge in the modeling and would be necessary for a robust segmentation.

Image resolution also affects the segmentation performance. The test data for healthy spine has two reconstructions of slice thickness: 0.7mm and 1mm. The best performer (Method 4) achieved 96.3% for 0.7mm data sets and 95.7% for 1mm data sets respectively. The diseased data set has 1mm and 2mm reconstructed slice thicknesses. The best performer (Method 1) achieved 90.6% for 1mm and 89.3% for 2mm data sets respectively. Intuitively, the segmentation algorithms perform better on higher resolution data set. New multi-channel CT scanners generate high resolution data. 1mm reconstructed CT is becoming a norm.

Model fitting or image registration is widely accepted as the reliable way to segment complex objects such as a vertebra. Different frameworks for the registration or optimization had been adopted by the participating methods. They all converged to a solution, however, at different rates and computational costs.

The running time ranged from a few minutes to 30 minutes (including the initialization), and three methods required manual initialization of the model. Vertebra segmentation is mostly needed for pre-operative planning, biomechanical simulation or offline diagnosis. Therefore, 30 minute segmentation time and manual interaction are accepted for clinical uses.

Since the vertebral bone has a relatively high contrast relative to its surrounding tissues, edge and gradient based feature functions were used in all methods. These feature functions could be sensitive to noise and compromised by surrounding bony tissues (e.g. ribs). Recently, however, machine learning techniques have been explored to classify pixels based on structural information and contextual features, which may mitigate the effect of this image noise. Furthermore, the gradients vary within the images due to variation of bone density and partial volume effect. The feature function must be adaptive to local image properties.

Method 1 and Method 4 have comparably the best performance in this study. Both methods include a component for automatic vertebra detection and localization, which makes them a more complete system. The initialization appeared to be rather robust since both methods successfully segmented all vertebrae. Method 4 performed best on the healthy cases and Method 1 performed best on the osteoporotic cases, although the difference was not statistically significant for the osteoporotic cases. Although different data sets were used in the evaluation, Method 1 and Method 2 outperformed the methods listed in Table 1 in terms of DICE coefficient and average surface distance.

There were a few limitations in this study report. Firstly, the data set was relatively small. However, it was very time consuming to generate the reference segmentation, especially for disease cases. Although we only have 20 data sets, each data set has 17 vertebrae. Therefore, each algorithm was tested on 340 vertebrae, which was a relatively large number. Secondly, the number of participants is relatively small. Computational spine imaging is a relatively small research field. Since this is the first comparative study in this field, we only managed to recruit five participants. This is actually the typical number of participants in most medical image analysis comparative studies. We have made the data public so that other researchers can test their algorithms. We also plan to keep the study open on spineWeb (http://spineweb.digitalimaginggroup.ca/spineweb). Thirdly, the reference standard is somewhat biased toward the result of the automatic segmentation that was used as the initial segmentation. Through the inter- and intra- observer experiments, the variability between two sets of manual segmentations is much lower than the difference between manual and computer segmentations. Even though the vertebral shape is complex, its border has distinct contrast to be located positively by an operator. Therefore, the

reference segmentation is consistent. Fourth, the data used in the comparative study were spine CT with intravenous contrast. From our experience, the vertebra segmentation algorithms perform better on non-contrast scans since less interference from contrast agent inside nearby aorta. In healthy spines, the interference is small since bone has much higher density than the contrast agent. However, the contrast agent may cause problems in osteoporotic cases since part of the spine may have similar density as the contrast agent. The contrast may be present inside the vertebral body where it will change the image voxel density and thus affect segmentation, density measurement and identification. A robust initialization and shape constrained vertebra model will provide a means to handle this issue. From the results of the participating algorithms, we didn't notice visible leakage into the aorta, which indicates that contrast had little impact on the vertebra segmentation algorithms. However, in this challenge, we didn't compare the performance between contrast studies and non-contrast studies.

The basic mechanism of the vertebra segmentation algorithms presented in this comparative study has potentials for application to more generalized clinical CT data sets. Through the comparison of various algorithms, readers should get a sense of the strengths and weaknesses of different vertebra segmentation algorithms and choose the appropriate one for their applications.

Acknowledgement

We thank the Department of Radiological Sciences, University of California, Irvine, Medical Center (Orange, CA, USA) for providing the CT data set. We thank Dr. Sasha Getty, Mr. Yinong Wang, and Mr. James Stieger for providing the manual segmentation for the reference data. We also thank the Imaging Biomarkers and Computer-Aided Diagnosis Laboratory, Radiology and Imaging Sciences Department, NIH (Bethesda, MD, USA) for providing the resource to carry out the evaluation. We also thank SpineWeb established by Digital Imaging Group of London (London, ON, Canada) for hosting the publicly available data set.

Figure Captions

Figure 1. Example of test cases with reference segmentation
Each vertebra is assigned a unique label (color coded). Sagittal and 3D views are shown.
Left: a healthy spine (test case 1)
Right: an osteoporotic and fractured spine (test case 10). Arrows point to vertebrae with compression fractures.

Figure 2. Partitioning of a vertebra into four substructures

Left: Density map on vertebra surface, hotter color: higher density
Right: Partitioning a vertebra into four substructures. The substructures are color-coded with different colors. The cutting planes lie at the border between two substructures.

Figure 3. Visual comparison of segmentation results for test case 2 (a healthy case)

Figure 4. Visual comparison of segmentation results for test case 10 (an osteoporotic and fractured case).

Figure 5. Visual comparison of segmentation results for specific vertebrae in test case 4
Row 1: T3 vertebra; Row 2: T9 vertebra; Row 3: L3 vertebra
Mid-axial slice for each vertebra is shown. The segmentation is superimposed on the CT data.

Figure 6. Mean performance of all methods for healthy and osteoporotic cases.
Both Dice Coefficient and mean surface distance are evaluated. The spine column is divided into three segments (T1-T6, T7-T12, and L1-L5) for assessment.

Figure 7. Mean performance of all methods for vertebra substructures.
The results show that vertebral body is more accurately segmented than other posterior sub structures.

Figure 8. Performance comparison on each vertebra group on healthy (left) and osteoporotic (right) cases. The charts compare the performance of different methods on each segment of the spine.

Figure 9. Performance comparison on the entire spinal column
Only lumbar vertebrae were evaluated in Method 5. The charts compare the difference of each method on healthy cases and osteoporotic cases.

Figure 10. Performance comparison on fractured and non-fractured vertebrae.
All methods except Method 2 performed better on non-fractured vertebrae.

Figure 11. Performance comparison on substructures.
The charts show the similar pattern of performance on substructures.

References

1. Burns, J.E., J. Yao, H. Muñoz, and R.M. Summers, *Automated Detection, Localization, and Classification of Traumatic Vertebral Body Fractures in the Thoracic and Lumbar Spine at CT.* Radiology, 2015. **142346**.
2. Iyer, S., B.A. Christiansen, B.J. Roberts, M.J. Valentine, R.K. Manoharan, and M.L. Bouxsein, *A Biomechanical Model for Estimating Loads on Thoracic and Lumbar Vertebrae.* Clin Biomech, 2010. **25**(9): p. 853-858.
3. Bourgeois, A.C., A.R. Faulkner, A.S. Pasciak, and Y.C. Bradley, *The evolution of image-guided lumbosacral spine surgery.* Annals of Translational Medicine, 2015. **3**(5): p. 69.
4. Kang, Y., K. Engelke, and W.A. Kalender, *A New Accurate and Precise 3-D Segmentation Method for Skeletal Structures in Volumetric CT Data.* IEEE TRANSACTIONS ON MEDICAL IMAGING, 2003. **22**(5): p. 586-598.
5. Li, H. and Z. Wang. *A Seepage Flow Model for Vertebra CT Image Segmentation*. in *IEEE Engineering in Medicine and Biology 27th Annual Conference*. 2005. Shanghai, China.
6. Aslan, M.S., A. Ali, H. Rara, and A.A. Farag. *AN AUTOMATED VERTEBRA IDENTIFICATION AND SEGMENTATION IN CT IMAGES*. in *IEEE 17th International Conference on Image Processing*. 2010.
7. Lim, P.H., U. Bagci, and L. Bai, *Introducing Willmore flow into level set segmentation of spinal vertebrae.* IEEE Transactions on Biomedical Engineering, 2013. **60**(1): p. 115-122.

8. Huang, J., F. Jian, H. Wu, and H. Li, *An improved level set method for vertebra CT image segmentation.* Biomedical engineering online, 2013. **12**(1): p. 48.
9. Li, Y., W. Liang, J. Tan, and Y. Zhang. *A novel automatically initialized level set approach based on region correlation for lumbar vertebrae CT image segmentation*. in *IEEE International Symposium on Medical Measurements and Applications* 2015.
10. Blumfield, A. and E. Blumfield, *AUTOMATED VERTEBRAL BODY IMAGE SEGMENTATION FOR MEDICAL SCREENING*, 2014: US.
11. Yao, J., S.D. O'Connor, and R.M. Summers. *Automated spinal column extraction and partitioning*. in *Biomedical Imaging: Nano to Macro, 2006. 3rd IEEE International Symposium on*. 2006.
12. Naegel, B., *Using mathematical morphology for the anatomical labeling of vertebrae from 3-D CT-scan images.* Comput. Med. Imaging Grap., 2007. **31**(3): p. 141-156.
13. Mastmeyer, A., K. Engelke, C. Fuchs, and W. Kalender, *A hierarchical 3D segmentation method and the definition of vertebral body coordinate systems for QCT of the lumbar spine.* Medical Image Analysis, 2006. **10**(4): p. 560-577.
14. Burnett, S., G. Starkschall, C.W. Stevens, and Z. Liao, *A deformable-model approach to semi-automatic segmentation of CT images demonstrated by application to the spinal canal.* Med. Phys., 2004. **31**(2): p. 251-263.
15. Klinder, T., J. Ostermann, M. Ehm, A. Franz, R. Kneser, and C. Lorenz, *Automated model-based vertebra detection, identification, and segmentation in CT images.* Medical Image Analysis, 2009. **13**(3): p. 471-482.
16. Ma, J., L. Lu, Y. Zhan, X. Zhou, M. Salganicoff, and A. Krishnan. *Hierarchical segmentation and identification of thoracic vertebra using learning-based edge detection and coarse-to-fine deformable model*. in *Medical Image Computing and Computer-Assisted Intervention*. 2010. Springer.
17. Rasoulian, A., R. Rohling, and P. Abolmaesumi, *Lumbar spine segmentation using a statistical multi-vertebrae anatomical shape+pose model.* IEEE Transactions on Medical Imaging, 2013. **21**(10): p. 1890-1900.
18. Kim, Y. and D. Kim, *A fully automatic vertebra segmentation method using 3D deformable fences.* Comp. Med. Imag. and Graph, 2009. **33**(5): p. 343-352.
19. Kadoury, S., H. Labelle, and N. Paragios, *Automatic inference of articulated spine models in CT images using high-order Markov Random Fields.* Medical Image Analysis, 2011. **15**(4): p. 426-437.
20. Ibragimov, B., B. Likar, F. Pernus, and T. Vrtovec, *Shape Representation for Efficient Landmark-Based Segmentation in 3-D.* IEEE Transactions on Medical Imaging, 2014. **33**(4): p. 861-874.
21. Roberts, M.G., T.F. Cootes, and J.E. Adams. *Segmentation of lumbar vertebrae via part-based graphs and active appearance models*. in *MICCAI*. 2009.
22. Stern, D., B. Likar, F. Pernu, and T. Vrtovec, *Parametric modelling and segmentation of vertebral bodies in 3D CT and MR spine images.* Phys. Med. Biol. , 2011. **56**: p. 7505-7522.
23. Wang, Y., J. Yao, H. Roth, J.E. Burns, and R.M. Summers. *Multi-Atlas Segmentation with Joint Label Fusion of Osteoporotic Vertebral Compression Fractures on CT*. in *3rd MICCAI 2015 Workshop & Challenge on Computational Methods and Clinical Applications for Spine Imaging*. 2015.
24. Ghebreab, S. and A. Smeulders, *Combining strings and necklaces for interactive three-dimensional segmentation of spinal images using an integral deformable spine model.* IEEE Trans. Biomed. Eng., 2004. **51**(10): p. 1821-1829.
25. Huang, S.-H., Y.-H. Chu, S.-H. Lai, and C.L. Novak, *Learning-Based Vertebra Detection and Iterative Normalized-Cut Segmentation for Spinal MRI.* IEEE TRANSACTIONS ON MEDICAL IMAGING, 2009. **28**(10): p. 1595-1605.

26. Suzani, A., A. Rasoulian, A. Seitel, S. Fels, R.N. Rohling, and P. Abolmaesumi. *Deep Learning for Automatic Localization, Identication, and Segmentation of Vertebral Bodies in Volumetric MR Images*. in *SPIE Medical Imaging: Image-Guided Procedures, Robotic Interventions, and Modeling*. 2015.
27. Mirzaalian, H., M. Wels, T. Heimann, B.M. Kelm, and M. Suehling. *Fast and Robust 3D Vertebra Segmentation using Statistical Shape Models*. in *International Conference of the IEEE EMBS*. 2013. Osaka, Japan.
28. Ibanez, L. and W. Schroeder, *ITK Software Guide*2003: Kitware Inc.
29. Forsberg, D. *Atlas-Based Segmentation of the Thoracic and Lumbar Vertebrae*. in *2nd MICCAI Workshops on Computational Methods and Clinical Applications for Spine Imaging (CSI2014)*. 2014. Boston, USA: Springer.
30. Seitel, A., A. Rasoulian, R. Rohling, and P. Abolmaesumi. *Lumbar and thoracic spine segmentation using a statistical multi-object shape+pose model*. in *2nd MICCAI Workshops on Computational Methods and Clinical Applications for Spine Imaging (CSI2014)*. 2014. Boston, USA: Springer.
31. Hammernik, K., T. Ebner, D. Stern, M. Urschler, and T. Pock. *Vertebrae Segmentation in 3D CT Images based on a Variational Framework*. in *2nd MICCAI Workshops on Computational Methods and Clinical Applications for Spine Imaging (CSI2014)*. 2014. Boston, USA: Springer.
32. Korez, R., B. Ibragimov, B. Likar, F. Pernus, and T. Vrtovec. *Interpolation-Based Shape-Constrained Deformable Model Approach for Segmentation of Vertebrae from CT Spine Images*. in *2nd MICCAI Workshops on Computational Methods and Clinical Applications for Spine Imaging (CSI2014)*. 2014. Boston, USA: Springer.
33. Castro-Mateos, I., J.M. Pozo, and A.F. Frangi. *3D Vertebra segmentation by feature selection Active Shape Model*. in *2nd MICCAI Workshops on Computational Methods and Clinical Applications for Spine Imaging (CSI2014)*. 2014. Boston, USA: Springer.
34. Dice, L.R., *Measures of the amount of ecologic association between species.* Ecology, 1945. **26**: p. 297-302.
35. Yao, J., J. Burns, S. Getty, J. Stieger, and R. Summers. *Automated extraction of anatomic landmarks on vertebrae based on anatomic knowledge and geometrical constraints*. in *International Symposium on Biomedical Imaging*. 2014. Beijing, China.
36. Forsberg, D., C. Lundstrom, M. Andersson, L. Vavruch, H. Tropp, and H. Knutsson, *Fully automatic measurements of axial vertebral rotation for assessment of spinal deformity in idiopathic scoliosis.* Physics in medicine and biology 2013. **58**(6): p. 1775-87.
37. Glocker, B., J. Feulner, A. Criminisi, D.R. Haynor, and E. Konukoglu, *Automatic Localization and Identification of Vertebrae in Arbitrary Field-of-View CT Scans.* Medical Image Computing and Computer-Assisted Intervention, 2012. **7512**: p. 590-598.
38. Ibragimov, B., R. Korez, B. Likar, F. Pernus, and T. Vrtovec. *Interpolation-Based Detection of Lumbar Vertebrae in CT Spine Images*. in *2nd MICCAI Workshops on Computational Methods and Clinical Applications for Spine Imaging (CSI2014)*. 2014. Boston, USA: Springer.