Deep Learning for Automatic Localization, Identification, and Segmentation of Vertebral Bodies in Volumetric MR Images

Amin Suzani[a], Abtin Rasoulian[a], Alexander Seitel[a], Sidney Fels[a], Robert N. Rohling[a,b], and Purang Abolmaesumi[a]

[a] Department of Electrical and Computer Engineering, University of British Columbia, Vancouver, BC, CANADA

[b] Department of Mechanical Engineering, University of British Columbia, Vancouver, BC, CANADA

ABSTRACT

This paper proposes an automatic method for vertebra localization, labeling, and segmentation in multi-slice Magnetic Resonance (MR) images. Prior work in this area on MR images mostly requires user interaction while our method is fully automatic. Cubic intensity-based features are extracted from image voxels. A deep learning approach is used for simultaneous localization and identification of vertebrae. The localized points are refined by local thresholding in the region of the detected vertebral column. Thereafter, a statistical multi-vertebrae model is initialized on the localized vertebrae. An iterative Expectation Maximization technique is used to register the vertebral body of the model to the image edges and obtain a segmentation of the lumbar vertebral bodies. The method is evaluated by applying to nine volumetric MR images of the spine. The results demonstrate 100% vertebra identification and a mean surface error of below 2.8 mm for 3D segmentation. Computation time is less than three minutes per high-resolution volumetric image.

Keywords: Vertebra localization, vertebra identification, spine segmentation, deep learning, statistical model registration.

1. INTRODUCTION

Localization and labeling of vertebrae is a crucial step in diagnosis and therapy of spinal diseases such as slipped vertebra, disk/vertebra degeneration, and spinal stenosis. Vertebra segmentation is also a pre-processing step in diagnosis of spine pathologies like scoliosis and osteoporosis. Hence, building a computer-based system for spine diagnosis and therapy requires automatic localization, labeling and segmentation of vertebral structures.

Lower back pain, one of the most common types of pain, is usually caused by problems in the lumbar region of the spine. Lumbar vertebrae are mostly viewed by X-rays, Magnetic Resonance (MR) imaging, or Computed Tomography (CT). Among these, MR imaging shows the soft tissue better and also does not expose the patient to ionizing radiations. This has led to increased interest in MR technologies for imaging the spine in recent years. Automatic labeling and segmentation of vertebral bodies in MR images is challenging because of: 1) variation among images in terms of field-of-view, 2) repetitive nature of vertebral column, 3) lower contrast between bony structures and soft tissue compared to CT, 4) larger inter-slice gap in clinical MR images compared to CT, and 5) presence of field inhomogeneities in MR images.

Several researchers have investigated the localization and labeling problem in CT[1–7] and MR[8–11] images. Most methods make assumption about which vertebrae are visible in the scan. Alomari et al.[11] assumed approximate alignment between scans. Klinder et al.[5], Glocker et al.[3,4], and Rasoulian et al.[12] claimed handling general scans. However, all of these studies are done on CT scans and not MR images. In addition, Glocker et al. only provide labeling and do not provide segmentation. Klinder et. al. and Rasoulian et al. provide segmenation, but their algorithms have high computation time for arbitrary-field-of-view scans.

In our previous work[13], we proposed a semi-automatic segmentation approach for MR images based on a multi-vertebrae statistical shape+pose model. It required one user click for each vertebra to be segmented. In this work, we propose a method based on deep learning for simultaneous localization and identification of vertebral bodies in MR images. Then, we initialize our previous segmentation method with the localized points in order to obtain a fully-automatic segmentation.

2. MATERIALS

The proposed method was evaluated on nine T1-weighted multi-slice MR images. All images contain lumbar region, but the extent of visibility of thoracic and sacrum regions varies. Slice thickness ranges between 3.3 to 4.4 mm among images. The size of all slices are 512×512 pixels with pixel spacing of 0.5 mm. The number of slices in each volumetric image varies from 12 to 18. The statistical model (used for segmentation) was constructed from manually segmented multi-slice CT scans of 32 patients[14]. Segmentation ground truth meshes were prepared manually using ITK-SNAP[15]. The center of gravity of each manually segmented vertebral body is used as a localization landmark.

3. METHODS

3.1 Automatic Localization and Identification

3.1.1 Pre-processing: Bias Field Correction

The presence of intensity inhomogeneities in MR images can adversely influence the quality of intensity-based feature extraction. Hence, we first apply a bias field correction algorithm on MR images to reduce this inhomogeneity. Statistical Parametric Mapping package is used for bias field correction (SPM12b, Wellcome Department of Cognitive Neurology, London, UK).

3.1.2 Localization and Identification by Deep Learning

The localization task is parametrized as a multi-variate regression problem. Each voxel of the image is described by hundreds of intensity-based features. Each feature is the difference between the mean intensity over two cuboids displaced with respect to the reference voxel position. Dimensions and displacement of each feature are generated randomly. These features are fast to compute using the integral image approach proposed by Viola and Jones[16]. The targets of the multi-variate regression are the relative distances between the reference voxel and the centers of lumbar vertebral bodies (Parametrization of image localization task as a regression problem, and also randomly generated cubic features are explained in more detail by Criminisi et al.[17]). A feed-forward neural network with three hidden layers is trained for solving the multi-variate regression problem. Stochastic gradient descent along with layerwise pre-training is used for optimization. On a test image, we first extract features from all voxels, then we use the neural network to predict the relative distances of the labels with respect to each voxel. Each relative distance is converted to the absolute position of the label and is considered as the vote of that specific voxel for the position of a specific vertebral body. The votes of all pixels are aggregated using a diffusion-based kernel density estimation[18] to obtain a robust prediction of the center of each lumbar vertebral body. No assumptions are made about the presence of the target vertebrae in the volumetric image. The voxels may vote outside of the scope of the image if the target vertebrae are not visible.

3.1.3 Refinement by Local Thresholding

In most of the cases, the predicted points from the deep learning approach are located inside of the target vertebral bodies, but not exactly centered. We attempt a simple approach to bring the predicted points to the center of vertebral bodies in order to improve the localization and also the initialization of our statistical model. By Otsu thresholding[19] in the region of the predicted points, we find the large components that may be vertebral bodies. If a large component is found close to the prediction, we refine the prediction by replacing it with the center of that component. Figure 1 illustrates an example of the refinement step.

3.2 Segmentation

3.2.1 Pre-processing: Anisotropic Diffusion

A three-dimensional anisotropic diffusion[20] filter is applied to the bias-field corrected image for image smoothing while preserving edges. This pre-processing step highly improves the quality of edge detection. For speed optimization, we only apply the filter to the region of spinal column detected by previous steps.

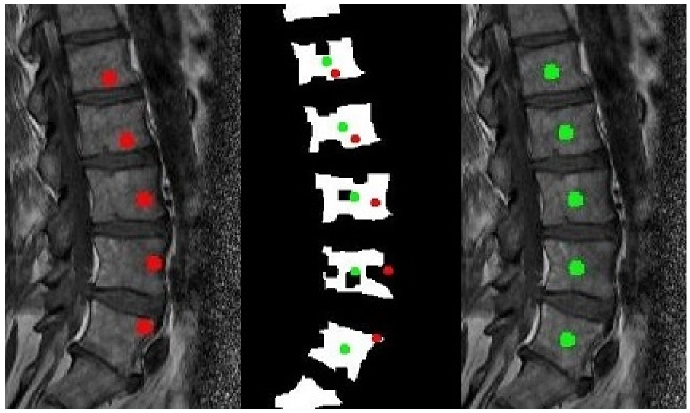

Figure 1: Refining localization points by replacing them with the center of the closest large component, obtained from local thresholding. Left: Localized points before refinement. Middle: Refinement using components obtained from local thresholding. Right: Localized points after refinement.

3.2.2 Statistical Model Registration

We use Canny edge detection algorithm[21] to obtain the edges in the detected spinal column region. The maximum value of the gradient magnitude of the region is used to automatically obtain the sensitivity threshold of the edge detector algorithm. Then, we use an iterative Expectation Maximization registration method (introduced by Rasoulian et al.[14]) to register a statistical model of vertebral bodies to Canny edges in order to obtain a robust segmentation. The statistical multi-vertebrae model integrates variations of shapes and poses of lumbar vertebrae that are separately extracted from 32 manually-segmented 3D CT scans[14]. The multi-vertebrae registration of the model allows us to avoid mis-segmentation in the area between vertebrae by simultaneous registration of all visible vertebral bodies and taking into account the correlation between shapes and poses of different vertebrae.

4. RESULTS AND DISCUSSION

Localization and identification results on nine subjects are reported in Table 1. The centroids of five lumbar vertebral bodies (L1 to L5) are predicted using leave-one-out cross validation. The distance to the ground truth landmarks are computed. The average and standard deviation of these distances as well as identification rate are reported. We consider a vertebral body to be correctly identified when its distance to the ground truth landmark is less than 2 mm and the closest landmark is the correct one. The same evaluation criteria is used by Glocker et al.[3] for identifying vertebrae in CT images. Our results show 100% correct identification after the refinement step for 45 lumbar vertebral bodies.

Three-dimensional segmentation results are shown in Table 2. The closest distance to the registered model is found for each point in the manual segmentation. The average and maximum of these distances are reported as mean error and Hausdorff distance respectively. These results demonstrate that we can automatically segment the lumbar vertebral bodies in volumetric MR images with a mean error of below 2.8 mm which shows improvement over our previous semi-automatic method[13]. Some examples of our identification and segmentation results are shown in Figures 2 and 3.

The localization and identification step (including refinement) takes less than 20 seconds on a desktop computer. The whole process takes less than 3 minutes using a MATLAB (The MathWorks, Inc., Natick, MA) implementation for the segmentation part. Training a deep neural network takes about one day which can be adjustable by downsampling the training data. Deep learning part is implemented in Python using Theano-nets package (Leif Johnson, Austin, TX) built on top of the Theano library[22].

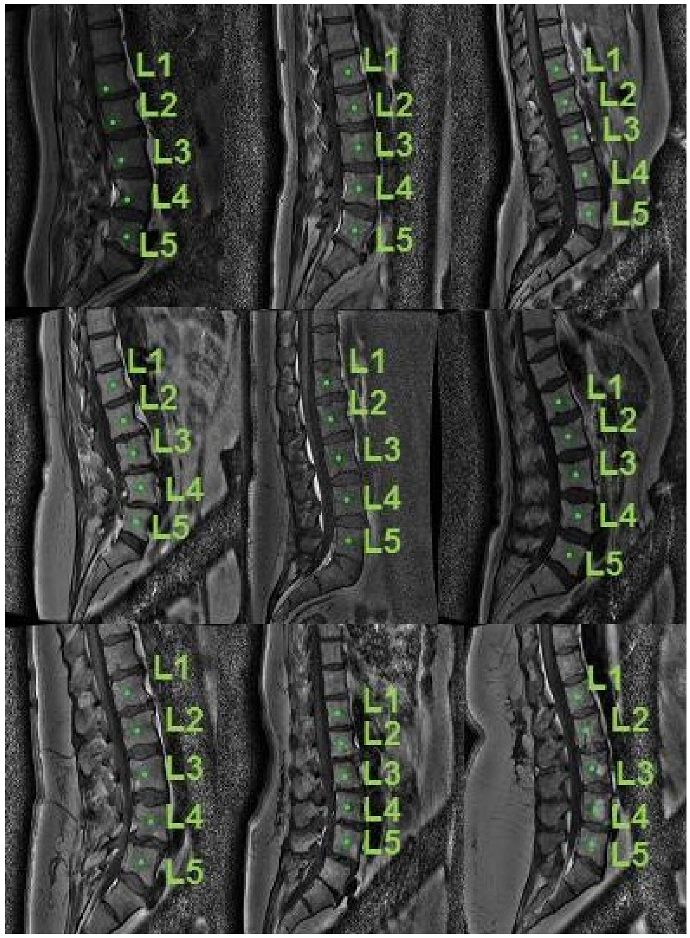

Figure 2: Localization and labeling results on mid-sagittal slice of bias-field corrected images.

Table 1: Localization and identification results for lumbar vertebral bodies in nine volumetric MR images.

	Mean error	Standard deviation	Identification
Deep Learning Localization	11.9 mm	6.3 mm	91%
After Refinement	3.0 mm	2.4 mm	100%

5. CONCLUSION AND FUTURE WORK

A fully-automatic approach based on deep learning and statistical models is proposed for localization, labeling and segmentation of vertebral bodies in multi-slice MR images. Although the experiments are performed only on the lumbar region, no assumptions are made about presence of specific vertebrae in the method. The multi-vertebrae model can handle large inter-slice gap in clinical MR images with low computation cost. Results demonstrate that our method can automatically localize, label, and segment the vertebral bodies in MR images with sufficient accuracy and speed for a wide range of clinical applications.

Future work will include an extensive evaluation of the method by using a more challenging dataset where different sections of the spine are visible in different images. The identification step also has the potential to work in real-time with a faster implementation of the feature extraction step.

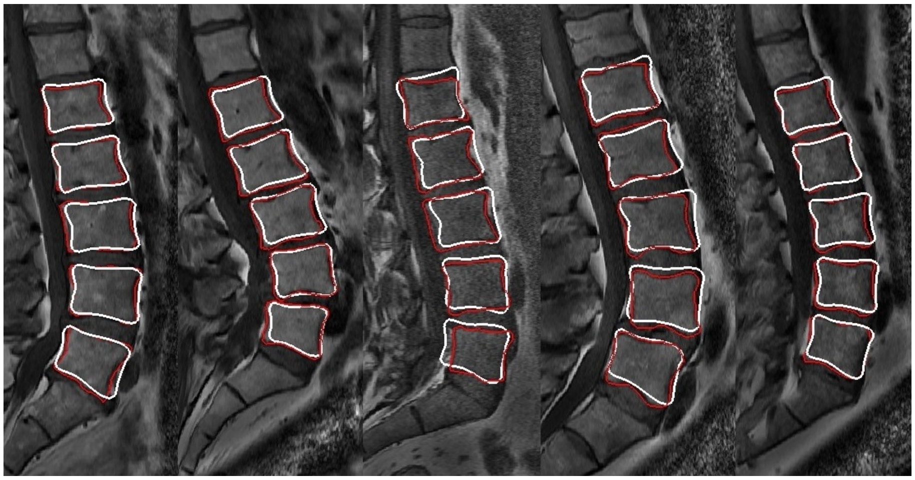

Figure 3: Examples of segmentation result in mid-sagittal slice of five different patients. Our segmentation results are shown with white contours, while red contours show the manual segmentation.

Table 2: Mean error and maximum distance (Hausdorff) of segmentation error (in mm) for each vertebral body.

Subject	L1		L2		L3		L4		L5		All lumbar vertebrae	
	Mean	Max	Mean	Max	Mean	Max	Mean	Max	Mean	Max	Mean $\pm$ std	Max
1	2.0	6.1	1.7	5.2	2.8	9.5	4.4	15.7	4.2	16.4	3.0 $\pm$ 1.2	16.4
2	2.2	6.5	2.2	8.0	2.0	7.3	1.8	9.6	3.4	16.6	2.3 $\pm$ 0.6	16.6
3	3.7	8.6	3.4	8.7	3.6	9.6	4.0	13.6	5.8	19.9	4.1 $\pm$ 1.0	19.9
4	2.6	8.1	1.9	7.1	2.6	10.5	1.6	8.3	2.0	9.1	2.1 $\pm$ 0.4	10.5
5	2.2	8.0	1.6	5.5	1.6	6.4	2.2	11.0	4.6	17.7	2.5 $\pm$ 1.2	17.7
6	2.7	10.3	2.6	9.7	2.8	9.3	2.5	10.4	4.0	16.1	2.9 $\pm$ 0.6	16.1
7	1.8	5.5	1.7	5.6	2.2	7.9	2.7	11.9	3.5	15.2	2.3 $\pm$ 0.7	15.2
8	3.2	8.4	3.3	9.5	3.2	10.4	3.6	12.7	2.6	11.0	3.2 $\pm$ 0.4	12.7
9	2.3	7.3	1.7	6.5	2.0	8.0	1.9	8.8	3.1	12.4	2.2 $\pm$ 0.5	12.4
Average	2.5	7.7	2.2	7.3	2.5	8.8	2.7	11.3	3.7	14.9	2.7 $\pm$ 0.9	19.9

6. ACKNOWLEDGEMENT

This work is funded in part by the Natural Sciences and Engineering Research Council of Canada (NSERC), and the Canadian Institutes of Health Research (CIHR). The three-dimensional MR images are generously provided by Dr. Shuo Li, GE Health Research, Ontario, Canada.

REFERENCES

[1] Ma, J., Lu, L., Zhan, Y., Zhou, X., Salganicoff, M., and Krishnan, A., "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–MICCAI 2010*], 19–27, Springer (2010).

[2] Hanaoka, S., Fritscher, K., Schuler, B., Masutani, Y., Hayashi, N., Ohtomo, K., and Schubert, R., "Whole vertebral bone segmentation method with a statistical intensity-shape model based approach," in [*SPIE medical imaging*], 796242–796242, International Society for Optics and Photonics (2011).

[3] Glocker, B., Feulner, J., Criminisi, A., Haynor, D. R., and Konukoglu, E., "Automatic localization and identification of vertebrae in arbitrary field-of-view CT scans," in [*Medical Image Computing and Computer-Assisted Intervention–MICCAI 2012*], 590–598, Springer (2012).

[4] Glocker, B., Zikic, D., Konukoglu, E., Haynor, D. R., and Criminisi, A., "Vertebrae localization in pathological spine CT via dense classification from sparse annotations," in [*Medical Image Computing and Computer-Assisted Intervention–MICCAI 2013*], 262–270, Springer (2013).

[5] Klinder, T., Ostermann, J., Ehm, M., Franz, A., Kneser, R., and Lorenz, C., "Automated model-based vertebra detection, identification, and segmentation in CT images," *Medical image analysis* **13**(3), 471–482 (2009).

[6] Naegel, B., "Using mathematical morphology for the anatomical labeling of vertebrae from 3D CT-scan images," *Computerized Medical Imaging and Graphics* **31**(3), 141–156 (2007).

[7] Rasoulian, A., Rohling, R., and Abolmaesumi, P., "A statistical multi-vertebrae shape+pose model for segmentation of CT images," in [*SPIE Medical Imaging*], **8671** (2013).

[8] Peng, Z., Zhong, J., Wee, W., and Lee, J.-h., "Automated vertebra detection and segmentation from the whole spine MR images," in [*Engineering in Medicine and Biology Society, 2005. IEEE-EMBS 2005. 27th Annual International Conference of the*], 2527–2530, IEEE (2006).

[9] Oktay, A. B. and Akgul, Y. S., "Simultaneous localization of lumbar vertebrae and intervertebral discs with SVM-based MRF," *Biomedical Engineering, IEEE Transactions on* **60**(9), 2375–2383 (2013).

[10] Schmidt, S., Kappes, J., Bergtholdt, M., Pekar, V., Dries, S., Bystrov, D., and Schnörr, C., "Spine detection and labeling using a parts-based graphical model," in [*Information Processing in Medical Imaging*], 122–133, Springer (2007).

[11] Alomari, R. S., Corso, J. J., and Chaudhary, V., "Labeling of lumbar discs using both pixel-and object-level features with a two-level probabilistic model," *Medical Imaging, IEEE Transactions on* **30**(1), 1–10 (2011).

[12] Rasoulian, A., Rohling, R. N., and Abolmaesumi, P., "Automatic labeling and segmentation of vertebrae in CT images," in [*SPIE Medical Imaging*], 903623–903623, International Society for Optics and Photonics (2014).

[13] Suzani, A., Rasoulian, A., Fels, S., Rohling, R. N., and Abolmaesumi, P., "Semi-automatic segmentation of vertebral bodies in volumetric MR images using a statistical shape+pose model," in [*SPIE Medical Imaging*], 90360P–90360P, International Society for Optics and Photonics (2014).

[14] Rasoulian, A., Rohling, R., and Abolmaesumi, P., "Lumbar spine segmentation using a statistical multi-vertebrae anatomical shape+pose model.," *IEEE Trans. Med. Imaging* **32**(10), 1890–1900 (2013).

[15] Yushkevich, P. A., Piven, J., Cody Hazlett, H., Gimpel Smith, R., Ho, S., Gee, J. C., and Gerig, G., "User-guided 3D active contour segmentation of anatomical structures: Significantly improved efficiency and reliability," *Neuroimage* **31**(3), 1116–1128 (2006).

[16] Viola, P. and Jones, M. J., "Robust real-time face detection," *International journal of computer vision* **57**(2), 137–154 (2004).

[17] Criminisi, A., Robertson, D., Pauly, O., Glocker, B., Konukoglu, E., Shotton, J., Mateus, D., Möller, A. M., Nekolla, S., and Navab, N., "Anatomy detection and localization in 3D medical images," in [*Decision Forests for Computer Vision and Medical Image Analysis*], 193–209, Springer (2013).

[18] Botev, Z., Grotowski, J., Kroese, D., et al., "Kernel density estimation via diffusion," *The Annals of Statistics* **38**(5), 2916–2957 (2010).

[19] Otsu, N., "A threshold selection method from gray-level histograms," *Automatica* **11**(285-296), 23–27 (1975).

[20] Perona, P. and Malik, J., "Scale-space and edge detection using anisotropic diffusion," *Pattern Analysis and Machine Intelligence, IEEE Transactions on* **12**(7), 629–639 (1990).

[21] Canny, J., "A computational approach to edge detection," *Pattern Analysis and Machine Intelligence, IEEE Transactions on* (6), 679–698 (1986).

[22] Bergstra, J., Breuleux, O., Bastien, F., Lamblin, P., Pascanu, R., Desjardins, G., Turian, J., Warde-Farley, D., and Bengio, Y., "Theano: a CPU and GPU math expression compiler," in [*Proceedings of the Python for Scientific Computing Conference (SciPy)*], (June 2010). Oral Presentation.