CrossMark

Detection and Labeling of Vertebrae in MR Images Using Deep Learning with Clinical Annotations as Training Data

Daniel Forsberg[1,2] · Erik Sjöblom[1] · Jeffrey L. Sunshine[2]

Abstract The purpose of this study was to investigate the potential of using clinically provided spine label annotations stored in a single institution image archive as training data for deep learning-based vertebral detection and labeling pipelines. Lumbar and cervical magnetic resonance imaging cases with annotated spine labels were identified and exported from an image archive. Two separate pipelines were configured and trained for lumbar and cervical cases respectively, using the same setup with convolutional neural networks for detection and parts-based graphical models to label the vertebrae. The detection sensitivity, precision and accuracy rates ranged between 99.1–99.8, 99.6–100, and 98.8–99.8% respectively, the average localization error ranges were 1.18–1.24 and 2.38–2.60 mm for cervical and lumbar cases respectively, and with a labeling accuracy of 96.0–97.0%. Failed labeling results typically involved failed S1 detections or missed vertebrae that were not fully visible on the image. These results show that clinically annotated image data from one image archive is sufficient to train a deep learning-based pipeline for accurate detection and labeling of MR images depicting the spine. Further, these results support using deep learning to assist radiologists in their work by providing highly accurate labels that only require rapid confirmation.

Keywords Archive · Artificial neural networks (ANNs) · Machine learning · Magnetic resonance imaging

✉ Daniel Forsberg
daniel.forsberg@sectra.com

[1] Sectra, Teknikringen 20, 583 30 Linköping, SE, Sweden

[2] Department of Radiology, Case Western Reserve University and University Hospitals Cleveland Medical Center, 11100 Euclid Avenue, Cleveland, OH 44106, USA

Introduction

Deep learning [1–3], a form of machine learning where neural networks with multiple hidden layers are trained to perform a certain task, has gained a widespread interest from a large number of disparate domains over the last few years, including radiology. Deep learning has repeatedly been shown to outperform other approaches. A popular deep learning application is object recognition [4, 5], i.e., when a network is trained to determine whether an image depicts a certain class of objects or not. Within radiology, deep learning has been shown useful for both text and image analytics [6–9].

A challenge associated with deep learning is the need for a large number of data samples to be used during the training phase. For example, the popular Mixed National Institute of Standards and Technology (MNIST) dataset used for training algorithms to classify handwritten digits contains 60,000 data samples for training, i.e., approximately 6000 samples per class. The need for large amounts of training data poses a challenge for deep learning within radiology, as large datasets with accurate annotations are rarely directly available. Hence, most deep learning-based projects within radiology will require a substantial manual effort to produce the data needed before training can commence. Consequently, access to relevant training data limits development and roll-out of applications based upon deep learning within radiology [10]. However, if one takes readily available clinical annotations from, for example, a picture archiving and communication system (PACS) as training data, then the required manual work would decrease substantially. One such set of available clinical annotations from a PACS are spine labels, i.e., text annotations providing a location and label of each vertebra visible in a spine image.

Most vertebrae in the human spine have a very similar shape (except C1, C2, and sacrum), especially when

🍎 Springer

comparing the vertebrae within each spine segment (cervical, thoracic, and lumbar) with each other. Consequently, a conscious identification of each vertebra is a necessity during the review of images depicting the spine to ensure that the correct levels of the spine are referenced when describing various findings. Further, since image studies of the spine typically consist of multiple image series, a radiologist will frequently manually annotate each vertebra with a label in a sagittal image to provide navigational support as the radiologist shifts focus among the different image series. This allows the radiologist to know the identity of all vertebrae without having to re-identify them each time another image series is viewed.

Detection and labeling of image data depicting the spine has already been considered by a number of research groups, both for computed tomography (CT) and magnetic resonance (MR). However, several of these have required training data in the form of fully segmented vertebrae [11–14], others with annotations in the form of bounding boxes [15–17] and just a few requiring only the position of each vertebra [18, 19]. Further, only some of the most recent works have considered deep learning [20–22] but none using clinically generated annotations as training data.

In this study, we investigate the feasibility of using clinically provided annotations of spine labels stored in a single PACS archive as training data for a deep learning-based pipeline capable of detection and labeling of vertebrae as depicted in sagittal MR images.

Materials and Methods

The conducted study was a retrospective study using already performed image studies. As such, the study was ruled exempt by the local institutional review board and informed consent was waived.

Image Data and Annotations

The local PACS was queried for spine labels of recent MR cases depicting the lumbar or the cervical spine. Once identified, associated image data and spine labels were exported

with subsequent HIPAA compliant de-identification. As the annotated spine labels were initially provided for navigational support, their placement was not always accurate or consistent. Hence, a manual quality assurance step was necessary to ensure that any retained cases would have all spine labels placed within the corresponding vertebrae and be consistently placed. Figure 1 contains examples of both retained and rejected cases. In total, 475 lumbar and 245 cervical cases were retained, containing 465/438 and 223/221 T1−/T2-weighted image series, and 3456 and 2321 spine labels, respectively. Note that not all cases contained both T1- and T2-weighted images and that the number of labeled vertebrae per case varied, although all contained at least L1-L5 and C2-C7 for the lumbar and cervical cases, respectively. Since the typical user scenario for spine labeling involves placing spine labels in a single image (typically the mid-slice within a sagittal image series), only the mid-sagittal slices from each image series were used. This design choice somewhat limits the applicability of the proposed pipelines, for example, in the case of scoliosis. However, in an actual implementation this can easily be handled as shown in [16]. The image data was resampled to a $1 \times 1 \text{ mm}^2$ resolution to ensure a uniform resolution across all images. Finally, the data was randomly split into three sets, training, validation, and test (60%/20%/20%), per spine region and MR sequence.

Detection and Labeling Pipeline

Two separate detection and labeling pipelines were constructed, one for the lumbar and one for the cervical cases. However, both pipelines were configured in the same manner, each with two convolutional neural networks (CNNs) for detection of potential vertebrae (one general lumbar/cervical vertebra detector and one specific S1/C2 vertebra detector, respectively) followed by a parts-based graphical model for false positive removal and subsequent labeling. The general vertebrae detectors would also detect any thoracic vertebrae that were depicted in the images. Figure 2 provides a schematic overview of the used pipeline for cervical cases.

All CNNs had similar network architecture with two convolutional (C) layers followed by a max-pooling (MP)
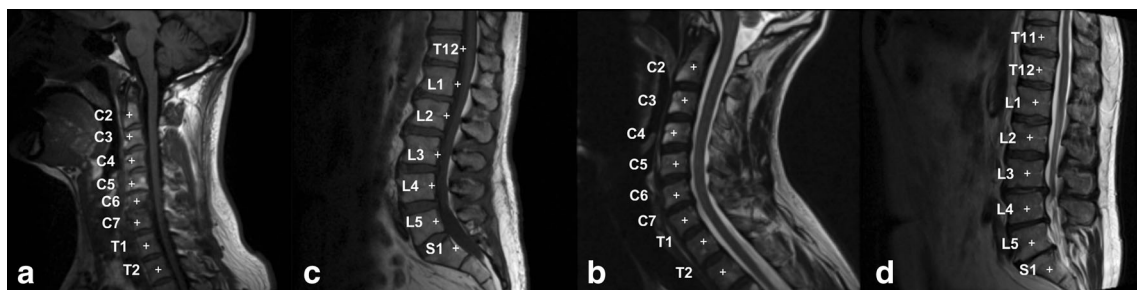

Fig. 1 Examples of cervical and lumbar cases with spine label annotations considered as inaccurate (label not correct or marker not in a centered position), (a) and (b), and accurate, (c) and (d), respectively
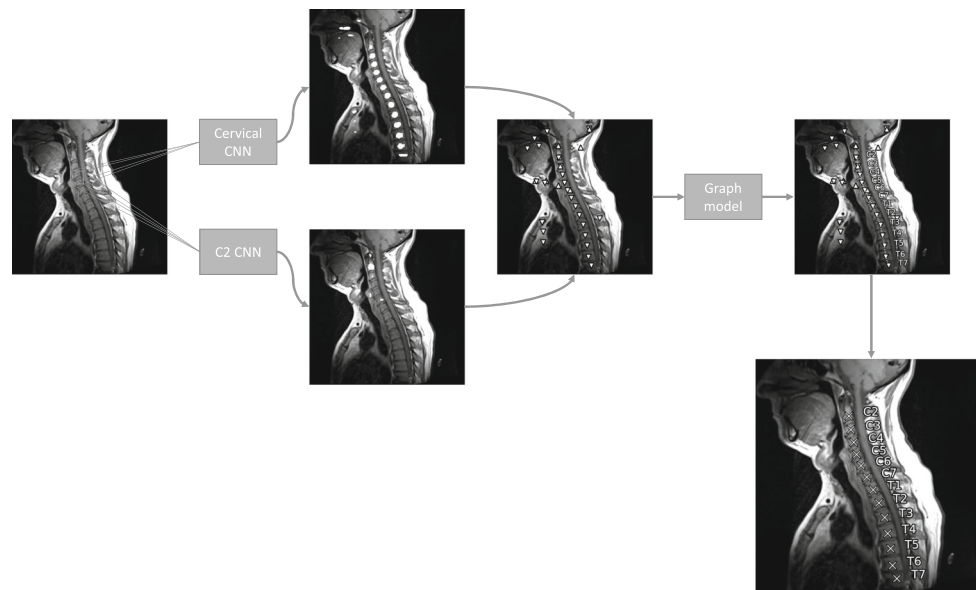
Fig. 2 Setup of the detection and labeling pipeline for cervical vertebrae. Image patches are extracted from the original image and fed to both the general cervical and the specific C2 CNNs, providing positive pixel-wise detections for image patches corresponding to a vertebra. After connected component analysis of the detection maps, the centroids of the detections are retained and fed to a parts-based graphical model, which discards false positive detections and labels the remaining detections. This provides the final output of a set of annotated spine labels

layer, another two C layers followed by a MP layer and finally, two fully connected (FC) layers before the subsequent output layer. All layers, except the output layer used a leaky rectify function as activation function and where the output layer used a soft-max function. The C and MP layers all used 3×3 receptive fields.

The size of the image patches and field-of-view differed somewhat between the CNNs, primarily depending on the approximate size of the relevant vertebrae. The patch sizes were 24×24 and 32×32 for the lumbar, respectively, the S1 CNNs, with a spatial resolution of 2×2 mm^2. The cervical and C2 CNNs had a patch size of 32×32 with a spatial resolution of 1×1 mm^2.

The CNNs would, unfortunately, provide a number of false positive detections in addition to any true positive detection, both for the general and the specific vertebra detectors. Hence, there was a need to find the most likely combination of detections among all the possible detections representing either the lumbar or the cervical spine. A parts-based graphical model, similar as in [16] (based upon [23]), was used for this purpose. The parts-based graphical model combines available detections and labels them to provide a configuration of detections and labels that best resembles earlier seen configurations of spine labels based upon label order and spatial locations.

All positive detections from each CNN were grouped using connected component analysis, and where the centroid and the area (normalized against the maximum area of all connected components for that particular image) of each connected component were returned as output from the detection step. Hence,

providing a location x_i and probability p_i for each vertebra detection i.

A layered (one layer for each expected vertebrae to label) graph $G = (V, E)$ with vertices $v_{i,l} \in V$ for every vertebra detection, $i \in (1, ..., N)$ and expected vertebrae to label, $l \in (1, ..., L)$ was constructed, with the layers in a stack according to the anatomical order of the vertebrae labels and only allowing edges between adjacent layers. The distance function of each edge $\delta(v_{i,l}, v_{j,l+1})$ was defined as $d_l(x_i, x_j) = (y_{i,j} - \mu_l)^T \Sigma^{-1} (y_{i,j} - \mu_l) \times 1/(p_i \times p_j)$ if $y_{i,j} \in R_l$ and 0 otherwise, where $y_{i,j} = x_j - x_i$. The variables μ_l, Σ_l, and R_l are inferred from the training data and correspond to the mean displacements ($\mu_1, ..., \mu_{L-1}$), the covariance matrices of the displacements ($\Sigma_1, ..., \Sigma_{L-1}$) and the allowed ranges for the displacements ($R_1, ..., R_{L-1}$) between every pair of adjacent vertebrae. Hence, finding the shortest path while including as many labels as possible, will provide the optimal configuration of vertebrae as given by all vertebra detections. The shortest path itself is easily found using e.g., Dijkstra's algorithm [23].

Training Data

Since the spine labels only correspond to a single point for each vertebra, there was a need to increase the number of image patches that could be used for training, both for positive (vertebra) and negative (non-vertebra) data samples. The pixels of each image, based upon the provided spine label locations, were coarsely split into four categories: vertebra center regions, intervertebral disk regions, vertebral edge

Fig. 3 Example of coarse pixel classification for cervical (**a**) and lumbar (**b**) cases, respectively. Pixels within the inner circular (*yellow*) regions are considered as vertebra center pixels, within the ellipsoid-like (*green*) regions as intervertebral disk pixels, within the outer circular (*red*) region as vertebra edge pixels and remaining as background pixels
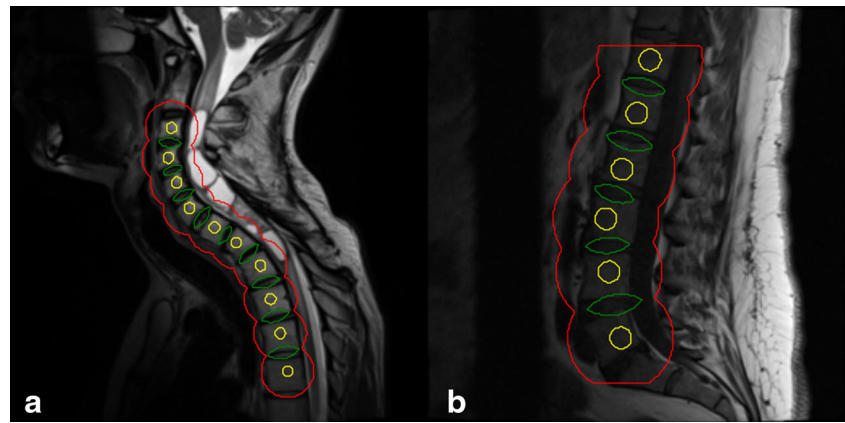
regions and background (Fig. 3). The vertebra center regions were determined as smaller circular regions around each spine label location; the intervertebral disk regions as the intersection of two circular regions centered in two adjacent vertebrae, respectively, with a radius corresponding to 65% of the distance between the two; the vertebral edge regions as larger circular regions around each spine label location but complementary to the vertebra center and intervertebral disk regions. Everything else was set as background.

For training and validation, all image patches corresponding to the vertebra center regions were kept as positive samples whereas image patches from the other regions were randomly selected to have the same number of samples as the positive samples from each of the three categories. Hence, in the end, the data consisted of approximately equal amounts of samples from all four categories. However, note that only data samples from the first category was used as (positive) vertebra samples whereas all other were used as (negative) non-vertebra samples, i.e. the data set was somewhat unbalanced. To avoid further unbalancing of the training data while at the same time maximizing the amount of training data was the initial reason for having two separate CNNs per pipeline, one for general vertebrae and one for specific vertebrae.

Training and Implementation Details

The training data was used for training the CNNs and building the parts-based graphical model per spine region and imaging sequence, whereas the validation data was used for finding a suitable CNN configuration along with determining optimal hyper parameters for the training. Both the CNNs and the parts-based graphical model were implemented in Python, where the CNNs were implemented using the freely available Lasagne library (https://github.com/Lasagne/Lasagne), which in turn is based upon the popular deep-learning library Theano [24]. During the training of the CNNs, the FC layers used 50% drop-out and the categorical cross-entropy was used a cost

function. A stochastic gradient descent approach with a Nestorov momentum was used to minimize the cost function. Note that the CNNs were all trained from scratch that is no transfer training was employed.

Evaluation

The two proposed vertebra detection and labeling pipelines were evaluated using the test data set and in terms of detection sensitivity, precision, accuracy, localization error and labeling accuracy. For the detection metrics a true positive vertebra detection corresponded to a detection located within a vertebra, a false positive detection to a detection located outside a vertebra and a false negative to any missed vertebrae, except the most superior or inferior vertebra for the lumbar and cervical spine sections, respectively, when not fully visible. Note that the detections refer to the final detections given as output from the complete pipeline and not the initial pixel-wise detections provided from the CNNs. Localization error was evaluated in the form of mean and standard deviation of the distances between the true positive detections and the centroids of the corresponding vertebrae. Given the inaccurate and inconsistent positioning of the clinically annotated spine labels, it was necessary to manually place these vertebra centroids (only used for evaluation purposes). For labeling accuracy, a true positive label corresponded to a detection located within a vertebra and with the correct label, and where all other labels were considered as false positives or false negatives.

Results

The numerical results from the performed evaluation are given in Table 1 and visualizations of some complete detection and labeling results are provided in Fig. 4. No noticeable differences in the detection results were observed between either lumbar or cervical cases or between T1 or T2 images. Failed

Table 1 Results from the evaluation of the two detection and labeling pipelines for lumbar and cervical MR cases, with results on a per vertebra/label level

Case type	Imaging sequence	Detection [%]			Localization error [mm]	Labeling accuracy [%]
		Sensitivity	Precision	Accuracy		
Lumbar	T1-weighted	99.7 (740/742)	99.9 (740/741)	99.6 (740/743)	2.38 ± 1.47	96.5% (716/742)
Lumbar	T2-weighted	99.7 (739/741)	99.6 (739/742)	99.3 (739/744)	2.60 ± 1.57	97.0% (719/741)
Cervical	T1-weighted	99.1 (421/425)	99.8 (421/422)	98.8 (421/426)	1.18 ± 0.81	96.0% (409/426)
Cervical	T2-weighted	99.8 (408/409)	100 (408/408)	99.8 (408/409)	1.24 ± 1.01	96.6% (394/408)

detections results were either associated with missed detections caused by pathology or by partly hidden vertebrae (for example when the scan orientation is not aligned with spine orientation and some vertebrae are not visible in the mid-sagittal image), or associated with failed S1 detections (either S1 was missed or S2 was mistaken as S1) for the lumbar cases. Examples of failed detections are provided in Fig. 4j–l. The localization error for the cervical cases, 1.18–1.24 mm, is about half the size for the lumbar cases, 2.38–2.60 mm. This is primarily explained by the substantially smaller cervical compared with the lumbar vertebrae, and thus, the variation in localization is bound to be smaller. The labeling accuracy of 96.0–97.0% is the expected given that incorrect labels are a direct consequence of failed detections. As such, a one-step shift of the labels, e.g., S1, L5, ..., L1, T12 → L5, L4, ..., T12, T11, would in most cases be sufficient to adjust the labels and is easily implemented in a PACS. Examples of this scenario are given in Fig. 4k, l.

For comparison purposes, we include Table 2, which provides details on the results reported by the earlier referenced works on spine detection and labeling.

Discussion

The results show that two successful detection and labeling pipelines for cervical and lumbar MR cases have been configured and trained based upon clinically annotated spine labels as training data. These results suggest that clinically annotated image data can indeed be used as training data for image analysis pipelines including deep learning. However, the clinically provided annotations needed a quality assurance step to ensure the accuracy, since as initially observed in this study; the annotations might not always be sufficient or consistently placed. The needed effort was minimal though and only required a few seconds per case. This can be compared to performing a full manual annotation, which required 5–15 s per case. Although the overall saved time in this scenario is modest, the potential benefit for more complex annotations is obvious. Note that the detection and labeling pipeline has only been applied to a single 2D image (the mid-sagittal) for each case and still very good results were achieved. Some of the missed detections likely could have been identified if multiple images were processed. These results show the potential for an increased usage of already existing image archives, where
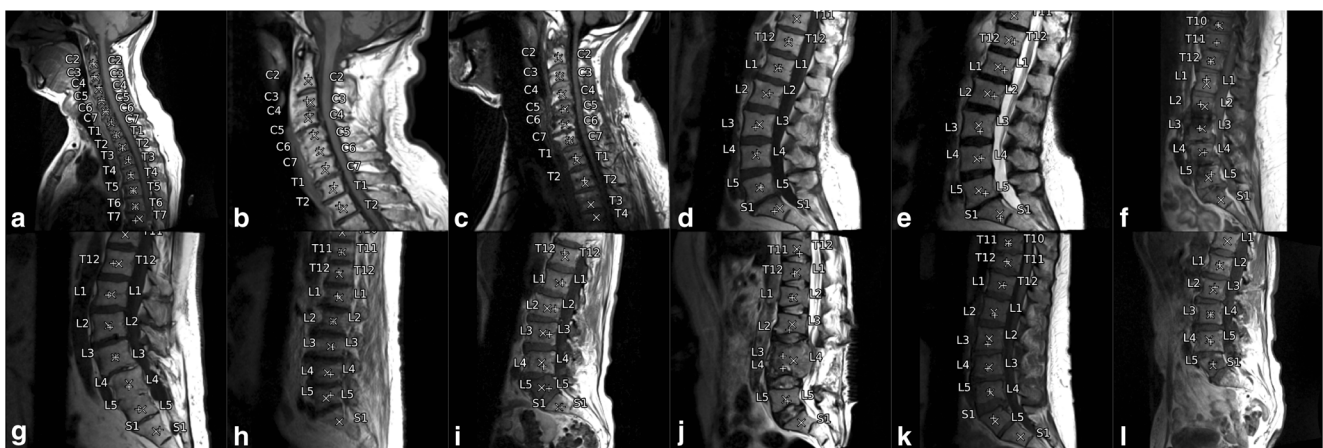

Fig. 4 Example results from the trained detection and labeling pipelines for cervical and lumbar MR images. Successful results are given in (**a**)–(**i**) and failed detections with subsequent failed labeling are given in (**j**)–(**l**).

Labels (+) on the anterior side correspond to clinically annotated spine labels and the labels (x) on the posterior side correspond to the spine labels as annotated by the trained pipelines

Table 2 Results reported by other research groups on their implemented methods for spine labeling

Work	Data	Detection rate [%]	Localization error [mm]	Labeling rate [%]
Huang et al. [15]	Cervical/lumbar MR	97.9/98.1	–	–
Klinder et al. [11]	Whole-spine and partial CT	92	–	95
Glocker et al. [18]	Whole-spine and partial CT	–	9.50	81
Glocker et al. [19]	Whole-spine and partial CT	–	7.0–14.3	62–86
Major et al. [13]	Disks in whole-spine CT	–	4.5	99.0
Oktay et al. [12]	Lumbar MR	–	2.95–3.25	97.8 (disks)
Lootus et al. [16]	Primarily lumbar MR	–	3.3	84.1
Cai et al. [17]	Whole-spine and partial CT/MR	–	2.08–3.44	93.8–98.2
Chen et al. [20]	Whole-spine and partial CT	–	8.82	84.16
Suzani et al. [21]	Whole-spine and partial CT	–	18.2	–
Zhan et al. [14]	Cervical/lumbar MR	98.0/98.8	3.07 (whole-spine)	–
Cai et al. [22]	Lumbar MR	–	3.23	98.1

archived images and associated annotations (typically rarely accessed) can provide additional value as a source of training data for deep learning.

Our presented results are on par with previously reported results, and surpass many of them, especially in measured localization errors. The localization errors vary substantially among the reported results with the largest localization errors for CT data sets covering the whole spine but also arbitrary partial sections of the spine [11, 18–21]. This comes as no surprise since the task of detection and labeling of arbitrary partial spine sections is much more difficult than the corresponding task for a specific complete spine section, e.g., the cervical or lumbar spine. For the localization error it should be noted though that our results are reported based upon 2D distances whereas some results in Table 2 are based upon 3D distances, which can be expected to be slightly larger.

An interesting comparison can be made between the clinically provided annotations and crowdsourcing. Crowdsourcing is often referred to as a process in which a crowd of people, usually online, is voluntarily engaged to either perform a certain task or to provide a monetary support for a project. This has for instance been used to provide large amounts of training data for deep learning [4]. Crowdsourcing has also been successfully employed within medical imaging to provide ground truth data [25]. As such, the clinically provided annotations can be considered as a type of crowdsourced training data, although provided by the radiologists as normal work effort. In other future learning situations, generation of training data could be established as part of the clinical routine work performed by the radiologists or other medical professionals. For example, if PACS vendors and radiologists together identify tasks that today are manually performed

by the radiologists but that would be possible to automate given sufficient available training data, then they could jointly devise a strategy to allow radiologists to establish the needed training data integrated in their routine work. Then over time and with little additional effort, a tool that fully or partially automates the previously manual tasks could be released later.

Another option to handle the issue of limited access to training data is to utilize transfer learning. In transfer learning, a pretrained CNN is trained on a new, but limited, data set where only a few of the last layers may be updated during training. Thus far, transfer learning has been shown to provide promising results [26, 27].

A limitation of the presented approach for detection and labeling is that in its current configuration it cannot handle cases where both the C2 or S1 vertebrae are not present, for example, as in cases with covering primarily the thoracic vertebrae. True though that this too is most difficult to obtain manually as well, and most often requires comparison with a larger field of view image obtained specifically to enable counting and labeling, that does include one or other end of the vertebral column. Alternatively, to handle these situations, additional CNNs are needed, trained for the detection of the C7/T1 and T12/L1 transitions, and most likely multiple 2D images need to be considered or even 3D image volumes that include ribs. Another limitation of the provided results is that the amount data used for the evaluation of the success of the automated vertebral detection and labeling was approximately 100 lumbar and 50 cervical cases. Further corroboration on data from more patients and from other institutions remains to prove further the robustness, consistency, and transferability.

References

1. Bengio Y.: Learning deep architectures for AI. Foundations and trends in Machine Learning 2(1):1–127,2009.

2. Bengio Y, Courville A, Vincent P: Representation learning: A review and new perspectives. IEEE Transactions on Pattern Analysis and Machine Intelligence 35(8):1798–1828,2013.

3. LeCun Y, Bengio Y, Hinton GE: Deep learning. Nature 521(7553): 436–444,2015.

4. Krizhevsky A, Sutskever I, Hinton GE: Imagenet classification with deep convolutional neural networks. In: Pereira F, Burges CJC, Bottou L, Weinberger KQ Eds. Advances in Neural Information Processing Systems,2012 25, 1097–1105

5. Szegedy C, Toshev A, Erhan D: Deep neural networks for object detection. In: Burges CJC, Bottou L, Welling M, Ghahramani Z, Weinberger KQ Eds. Advances in Neural Information Processing Systems,2013, 2553–2561

6. Shin HC, Lu L, Kim L, Seff A, Yao J, Summers RM: Interleaved text/image deep mining on a very large-scale radiology database. In: 2015 I.E. Conference on Computer Vision and Pattern Recognition (CVPR), 2015, 1090–1099

7. Pereira S, Pinto A, Alves V, Silva CA: Brain tumor segmentation using convolutional neural networks in MRI images. IEEE Transactions on Medical Imaging 35(5):1240–1251,2016.

8. Setio AAA, Ciompi F, Litjens G, et al.: Pulmonary nodule detection in CT images: false positive reduction using multi-view convolutional networks. IEEE Transactions on Medical Imaging 35(5):1160–1169,2016

9. Yan Z, Zhan Y, Peng Z, et al.: Multi-instance deep learning: Discover discriminative local anatomies for bodypart recognition. IEEE Transactions on Medical Imaging 35(5):1332–1343,2016

10. Greenspan H, van Ginneken B, Summers RM: Guest editorial deep learning in medical imaging: Overview and future promise of an exciting new technique. IEEE Transactions on Medical Imaging 35(5):1153–1159,2016

11. Klinder T, Ostermann J, Ehm M, Franz A, Kneser R, Lorenz C: Automated model-based vertebra detection, identification, and segmentation in CT images. Medical Image Analysis 13(3):471–482, 2009

12. Oktay AB and Akgul YS: Simultaneous localization of lumbar vertebrae and intervertebral discs with svm-based mrf. Biomedical Engineering, IEEE Transactions on 60(9):2375–2383,2013

13. Major D, Hladůvka J, Schulze F, Bühler K: Automated landmarking and labeling of fully and partially scanned spinal columns in CT images. Medical Image Analysis 17(8):1151–1163, 2013

14. Zhan Y, Jian B, Maneesh D, Zhou XS: Cross-Modality Vertebrae Localization and Labeling Using Learning- Based Approaches. In: Li S, Yao J Eds. Spinal Imaging and Image Analysis, 2015, 301–322

15. Huang SH, Chu YH, Lai SH, Novak CL: Learning-based vertebra detection and iterative normalized-cut segmentation for spinal MRI. Medical Imaging, IEEE Transactions on 2009 28(10):1595–1605, 2009.

16. Lootus M, Kadir T, Zisserman A: Vertebrae Detection and Labelling in Lumbar MR Images. In: Yao J, Klinder T, Li S Eds. Computational Methods and Clinical Applications for Spine Imaging, 2014, 219–230.

17. Cai Y, Osman S, Sharma M, Landis M, Li S: Multi-modality vertebra recognition in arbitrary views using 3d deformable hierarchical model. IEEE Transactions on Medical Imaging 34(8):1676–1693, 2015.

18. Glocker B, Feulner J, Criminisi A, Haynor DR, Konukoglu E. Automatic Localization and Identification of Vertebrae in Arbitrary Field-of-View CT Scans. In: Ayache N, Delingette H, Golland P, Mori K Eds. Medical Image Computing and Computer-Assisted Intervention – MICCAI 2012, 2012, 590–598

19. Glocker B, Zikic D, Konukoglu E, Haynor DR, Criminisi A: Vertebrae localization in pathological spine CT via dense classification from sparse annotations. In: Mori K, Sakuma I, Sato Y, Barillot C, Navab N Eds. Medical image computing and computer-assisted intervention – MICCAI 2013. Heidelberg: Springer Berlin, 2013, 262–270.

20. Chen H, Shen C, Qin J, Ni D, Shi L, Cheng JCY, Heng PA: Automatic Localization and Identification of Vertebrae in Spine CT via a Joint Learning Model with Deep Neural Networks. In: Navab N, Hornegger J, Wells WM, Frangi AF Eds. Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015, 2015, 515–522

21. Suzani A, Seitel A, Liu Y, Fels S, Rohling RN, Abolmaesumi P: Fast Automatic Vertebrae Detection and Localization in Pathological CT Scans - A Deep Learning Approach. In: Navab N, Hornegger J, Wells WM, Frangi AF Eds. Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015, 2015, 678–686

22. Cai Y, Landis M, Laidley DT, Kornecki A, Lum A, Li S: Multi-modal vertebrae recognition using transformed deep convolution network. Computerized Medical Imaging and Graphics 51:11–19, 2016.

23. Dijkstra EW: A note on two problems in connexion with graphs. Numerische Mathematik 1(1):269–271,1959.

24. Al-Rfou R, Alain, G, Almahairi A: Theano: A Python framework for fast computation of mathematical expressions. arXiv e-prints 2016;abs/1605.02688. http://arxiv.org/abs/1605.02688. Published May 9, 2016. Accessed June 6, 1026,2016

25. Volpi D, Sarhan MH, Ghotbi R, Navab N, Mateus D, Demirci S: Online tracking of interventional devices for endovascular aortic repair. International Journal of Computer Assisted Radiology and Surgery 10(6):773–781,2015.

26. Shin HC, Roth HR, Gao M, et al.: Deep convolutional neural networks for computer-aided detection: CNN architectures, dataset characteristics and transfer learning. Medical Imaging, IEEE Transactions on 35(5):1285–1298,2016.

27. Tajbakhsh N, Shin JY, Gurudu SR, Hurst RT, Kendall CB, Gotway MB, Liang J: Convolutional neural networks for medical image analysis: full training or fine tuning? IEEE Transactions on Medical Imaging 35(5):1299–1312,2016.