IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received April 9, 2021, accepted May 22, 2021, date of publication May 31, 2021, date of current version June 23, 2021.

Digital Object Identifier 10.1109/ACCESS.2021.3084949

Deep Learning-Based Automatic Segmentation for Reconstructing Vertebral Anatomy of Healthy Adolescents and Patients With Adolescent Idiopathic Scoliosis (AIS) Using MRI Data

M. ANTICO[1,2], J. P. LITTLE[1,2], H. JENNINGS[1,3], G. ASKIN[1,3], R. D. LABROM[1,3], D. FONTANAROSA[4], AND P. PIVONKA[1,2]

[1]Centre for Children's Health Research, Biomechanics and Spine Research Group, Institute of Health and Biomedical Innovation (IHBI), Queensland University of Technology, Brisbane, QLD 4101, Australia
[2]School of Mechanical, Medical and Process Engineering, Queensland University of Technology, Brisbane, QLD 4000, Australia
[3]Queensland Children's Hospital, Brisbane, QLD 4101, Australia
[4]School of Clinical Sciences, Queensland University of Technology, Brisbane, QLD 4000, Australia

Corresponding author: M. Antico (m2.antico@qut.edu.au)

This work involved human subjects or animals in its research. Approval of all ethical and experimental procedures and protocols was granted by the Queensland University of Technology under Application No. 1200000281, the Children's Health Queensland under Application No. SSA/14/QRCH/411, and the Mater Health Human Research Ethics Committees under Application No. 14/88/AM03.

ABSTRACT MRI is a non-ionising imaging modality that could be used as an alternative to Xray-based imaging methods to accurately assess the 3D morphology of the vertebral anatomy of scoliosis patients. However, a major caveat in utilising MRI is the significant amount of time required to manually segment the anatomy of interest. To overcome this limitation, we implemented a fully automatic method for the 3D segmentation of thoracic vertebrae, including vertebral body and posterior elements, of healthy adolescents and patients with Adolescent Idiopathic Scoliosis (AIS) using MRI data. 62 MRI scans were obtained from 3 healthy volunteers and 25 patients with AIS. A state-of-the-art deep-learning network for segmentation was trained using image patches of the apical vertebra (T7, T8, T9 or T10) extracted from 20 AIS patient MRIs. Ad-hoc data augmentation was adopted to represent the unlabeled vertebral levels in the dataset (T5-T6, T11-T12). The vertebral levels T5-T12 were then segmented for the remaining MRI datasets by feeding to the network the MRI patches generated by translating a window of fixed size and stride onto the MRI volume. The mean dice score coefficient for the AIS patient vertebral levels T5-T12 was of 87% $\pm$ 4.3%, which was comparable to the performance achieved by two experts. On average, 93% and 97% of the MRI segmented slices were considered clinically acceptable morphological reconstructions of AIS and healthy volunteer vertebrae, respectively. The proposed method can be considered as the first step towards more routine MRI-based imaging of AIS osseous deformities, reducing the cumulative exposure of young patients to ionising radiation.

INDEX TERMS Deep learning, CNN, segmentation, scoliosis, MRI, vertebral body, posterior process, AIS.

I. INTRODUCTION

Adolescent Idiopathic Scoliosis (AIS) is a progressive multi-dimensional deformity of the spine found in about 2-3% of adolescents [1]. It is characterised by atypical growth of the vertebral body in the coronal, sagittal and axial planes, as well as asymmetrical development of the pedicles. In clinical practice, monitoring this spinal pathology is a determinant for patient treatment management [2]. Conservative treatment to delay scoliosis progression is achieved via spinal orthoses/bracing, while, in the more severe cases, surgical correction may be required. AIS assessment and follow up is performed using radiographic imaging (XRay or bi-planar, EOS®), which provides only 2D information about the deformed vertebral morphology, and exposes young patients to cumulative ionising radiation throughout their healthcare [3].

While 2D plain radiography is the basis for diagnosis, follow-up and monitoring of patients, AIS vertebrae are characterised by a 3D deformity of both the vertebral body and

The associate editor coordinating the review of this manuscript and approving it for publication was Anubha Gupta.

the posterior elements when compared to a typical developing adolescent spine. The ability to examine this 3D morphology in detail is important clinically for surgeons when assessing pathological spinal anatomy, and also in biomechanical investigations, where a virtual representation of the 3D bone surface (or segmentation) can be used to simulate patient-specific mechanics. Computed tomography (CT) imaging provides high-quality 3D imaging of the deformed spinal anatomy and is used clinically to identify underlying pathology and guide surgical treatment [4]. However, also CT-derived exposure of patients to ionising radiation, with the associated lifetime attributable risk of morbidity, leaves paediatric and adolescent patients particularly vulnerable due to their young age [5]–[7].

MR imaging could represent an alternative to radiation-based imaging by providing an accurate, radiation-free method to examine this pathology and, eventually, volumetrically analyse AIS vertebral growth and deformity progression. Furthermore, this imaging modality could also be valuable to inform surgical planning and/or intra-operative guidance for spinal surgery, where the 3D MRI-based spine model can be registered to fiducial markers on the patient spine [8]. However, despite its potential, a significant caveat in utilising MRI is the time required for manual segmentation of bone and cartilage structures, and the accuracy achieved [8].

This study aims at overcoming these limitations by implementing an automatic segmentation method for the 3D reconstruction of the vertebral anatomy of both AIS patients and healthy adolescents from MRI data. To enable accurate spine analysis, we reconstructed for the first time the entire thoracic vertebra bony anatomy, including the vertebral body and the posterior process.

Previous studies for 3D vertebrae automatic segmentation focused on the vertebral body of the spine, excluding the posterior process, which has a more complex and irregular shape. Earlier developed approaches typically involved model-based techniques, using balloon forces [9], elastic finite element models [10] or combinations of statistical shape modelling and active shape models [11]–[14]. More recent methods include traditional machine learning (ML)-based algorithms and convolutional neural networks (CNNs). In the former category, Chu *et al.* [15] utilised a random forest-based method first to localise the vertebral bodies centres, and then to segment the vertebral bodies combining the appearance learned during algorithm training and the shape information derived from Parzen windows. An alternative approach was proposed by Gaonkar *et al.* [16], who developed a segmentation method robust to different MRI scanning protocols. Vertebral bodies were segmented using an ensemble of random forests, each trained using different relevant vertebral body features, obtained for varying parameters settings of the Felzenswalb superpixel algorithm. Fallah *et al.* [17] developed a similar approach training random forests through image features extracted from MRI images at different resolutions. Gawel *et al.* [18] used, instead, a cascade of Boosted classifier combined with an active appearance model. All these methods strongly rely on hand-crafted features, based on the morphology or the image intensity of the region of interest. For this reason, they may not be robust in more complex imaging scenarios, considering higher image variability, e.g. in the presence of pathology.

CNNs have the advantage of extracting the relevant image features automatically, thus without the need for manual feature engineering. Several deep-learning-based approaches were reported in the literature for the vertebral anatomy segmentation from CT, including classification-based methods [19] and 2D/3D CNNs, such as state-of-the-art algorithms (UNet [20]) [21]–[23] or their variants [24]. 2D UNets [25], [26] and hybrid approaches combining CNN with model-based segmentation or star convex graph cuts [27], [28] were detailed for MRI-based vertebral body automatic segmentation and achieved the highest performance for this task. The approach developed by Rak. *et al.* [20] used a simplified version of the UNet trained with 2D MRI image patches to extract a map of the vertebrae likelihood, followed by a graph cut formulation combining the likelihood maps extracted with star-convexity constraints. In contrast, Korez *et al.* [21] extracted the vertebrae likelihood using a 3D CNN and used a deformable model to adapt this pre-segmentation to the vertebrae boundaries. The main limitation of the CNN-based algorithms is the need for a large amount of labelled training data that needs to represent the desired model output. Furthermore, data variability increases significantly when spinal deformities (or the more complex anatomy of the posterior process) are considered, increasing the complexity of the problem.

Among the implemented approaches, only Guerroumiet al [25] addressed vertebrae segmentation for AIS patients. In their work, to limit the amount of required labelled data, the CNN was first trained on a publicly available dataset of healthy volunteers and subsequently fine-tuned utilising AIS patient data. However, as for the other ML and CNN methods, their approach required a labelled training dataset comprehensive of all the vertebral levels to be segmented. In this paper, we propose an alternative, data-efficient method to automatically segment the thoracic spine (T5-T12) using a state-of-the-art CNN trained with a labelled dataset limited to the apical vertebra (i.e. the most deviated laterally from the central sacral line [29]) of each AIS patient. Differently from [25], [26], model training was performed using MRI image patches corresponding to the ground-truth labels (comprising the vertebral body and the posterior process) and adopting ad-hoc data augmentation to represent the unlabeled vertebral levels in the dataset. The vertebral levels T5-T12 were then segmented by feeding to the network the image patches obtained by translating a window of fixed size and stride onto the MRI volume. Our approach thus exploits the similarity among the different vertebral levels treating them as a single class and compensates for their variability through ad-hoc data augmentation and MRI decomposition into multiple patches.

II. MATERIALS AND METHODS

A. DATASET OF AIS PATIENTS AND HEALTHY VOLUNTEERS

A database including sequential MRI scans from 25 AIS patients and 3 healthy (non-scoliotic) volunteers was used for this study (Table 1). Ethics approval for data collection was granted by the Queensland University of Technology (1200000281), the Children's Health Queensland (SSA/14/QRCH/411) and the Mater Health Human Research Ethics Committees (14/88/AM03). For each AIS patient, 3 to 6 MRI scans were acquired, at 3 to 20 months intervals (Table 1). Of these MRI datasets, at least two scans were used for each patient - initial presentation for tertiary spinal clinic care, and just prior to either spinal fusion surgery, or their final clinical follow-up appointment. For healthy participants, annual MRI scans were acquired up to the point of skeletal maturity (Table 1). The first MRI scan for the 3 healthy participants was utilised in the current study.

TABLE 1. AIS and healthy participants' demographics.

	AIS	Healthy
Number of Subjects	25	3
Age at First Scan (years)[mean ± SD]	12.4 ± 1.3	11.5 ± 1.0
Age at Final Scan (years) [mean ± SD]	14.7 ± 1.1	14.1 ± 1.0
Mean time between scans (years) [mean ± SD]	2.3 ± 1.2	2.6 ± 0.5
Mean Cobb Angle at First Scan (degrees) [mean ± SD]	29 ± 18	N/A

SD = standard deviation

B. MRI SCANNING PROTOCOL

The participants of this study were examined on a 3 Tesla (3T) MRI scanner (Philips Achieva 3.0T TX Dual Transit System), using a T1-weighted 3D gradient echo sequence (Voxel size 0.5 × 0.5 × 0.5mm, TR 5.9ms, TE 2.7ms, Flip angle 5 degrees). The scans were acquired with the participants in supine position, with the arms resting along the torso. The MRI field of view obtained comprised at least the thoracic vertebrae from T5 to T12.

C. GROUND-TRUTH LABEL GENERATION

Ground-truth labels were generated for the apical vertebra of AIS patients and vertebra T9 of non-scoliotic participants, using a software for image analysis (AMIRA 6.0.0, ThermoFisher-Scientific, USA). For AIS patients, the apical vertebra for each patient was first identified on the coronal plane. The apical vertebral level was T7, T8, T9 and T10 for 2, 12, 9 and 2 AIS patients, respectively. The MRI axial slices of the vertebra of interest were manually segmented using a brush tool, and the generated segmentation stack converted to a binary mask [30]. The sagittal and the coronal projections of the mask were also examined and adjusted if necessary to ensure all vertebral regions were correctly selected.

The vertebrae T5, T6, T11 and T12 were also segmented on the MRI scans of 1 or 2 AIS patients following the same procedure, to enable testing the algorithm performance on these additional vertebral levels. Table 2 reports a summary of the labelled vertebral levels for AIS and healthy participants.

The labels were generated, as part of a larger clinical MRI dataset, by two experts: a trained medical doctor (H.J.), who is an orthopaedic registrar, and a Biomedical Engineer, who has been trained in 3D reconstruction of clinical MRI data. These researchers performed all the manual segmentations of MRI datasets for individual vertebrae, and in both instances, their observer error in reliably reconstructing osseous anatomy from MRI was confirmed (see Section II-F-3). This process of manual segmentation to generate 3D reconstructions of osseous anatomy from MRI data is the standard approach used clinically, and both these researchers were chosen as they had considerable experience in this workflow.

TABLE 2. Labelled volumes for AIS and healthy participants.

	Vertebral level	Apical vertebra	Number of volumes	Number of volunteers
AIS	T5	-	1	1
	T6	-	2	2
	T7	✓	4	2
	T8	✓	29	12
	T9	✓	21	9
	T10	✓	4	2
	T11	-	2	2
	T12	-	2	2
Healthy	T9	-	3	3

D. CNN TRAINING

1) IMAGE AND LABEL PRE-PROCESSING

The AIS patients' labelled dataset of the apical vertebra was utilised for algorithm training. Pre-processing was applied to this dataset using MATLAB (version 9.3.0 [R2017 b], The Mathworks Inc. Natick, MA, USA). For each MRI volume, the subvolume corresponding to the ground-truth label was selected (Figure 1). Along the anterior-posterior and the cranial-caudal directions, the boundaries of each subvolume were the extremes of the minimum bounding box of the corresponding ground-truth label. Along the medial-lateral direction, a fixed boundary was selected for all the volunteers' subvolumes, by visually inspecting the MRIs and excluding the outermost parts that did not contain the vertebrae. More specifically, the image region included 261 pixels, selected starting from the 119th pixel from the most lateral side of the MRI volume.

a: ANATOMICAL PLANE SELECTION

Due to the morphology of the spine, the subvolumes generated contained part of the superior and part of the inferior adjacent vertebrae, for which labelling was not provided. To enable a proper training of the algorithm, the anatomical

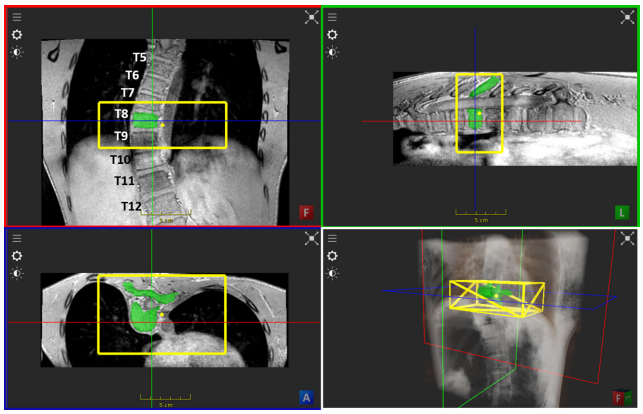

FIGURE 1. Example of MRI subvolume selection (yellow box) based on the corresponding ground-truth label (in green) shown for visualisation purposes in a software for image analysis (ImFusion, München, Germany). Red, green, blue and white panels in the figure correspond to the coronal, sagittal, axial and 3D rendering of the MRI, respectively.

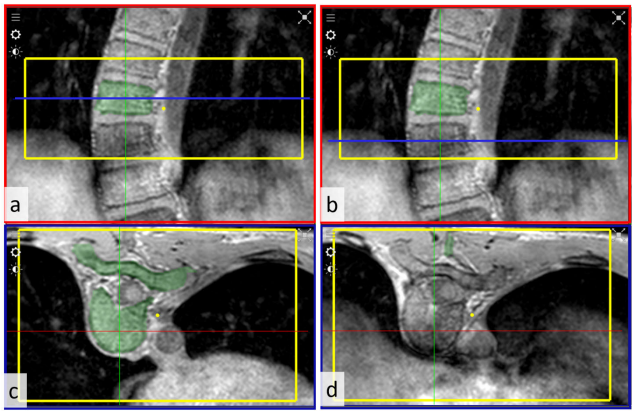

FIGURE 2. Representative example of MRI slices generated when the axial plane intersects the labelled (in green) and an unlabelled adjacent vertebra (a-c and b-d, respectively) within the MRI subvolume containing the labelled vertebra, enclosed in the yellow box. a)-b) show in blue the axial plane intersecting the vertebra and c)-d) the resulting axial MRI slices.

plane selected for slicing the subvolumes should not generate 2D images where the annotated and the not annotated (adjacent) vertebrae have similar shapes. The inconsistency in labelling would, in fact, not allow the algorithm to find a proper correlation between the images and the labels. As shown in Figure 2 a-b, this was the case for the axial plane.

Both the sagittal and the coronal plane were instead suitable candidates. The coronal plane was selected being the plane generating the highest number of 2D labelled images, as the longest dimension of the thoracic vertebrae was along the anterior-posterior direction (from the anterior vertebral body to the tip of the posterior process). The MRI subvolumes were thus sliced along this plane, and not-annotated slices were discarded from the dataset. This process resulted in 6449 2D MRI images of the apical vertebra with corresponding ground-truth labels. Figure 3 shows for each AIS patient the number of annotated coronal slices for each vertebral level.

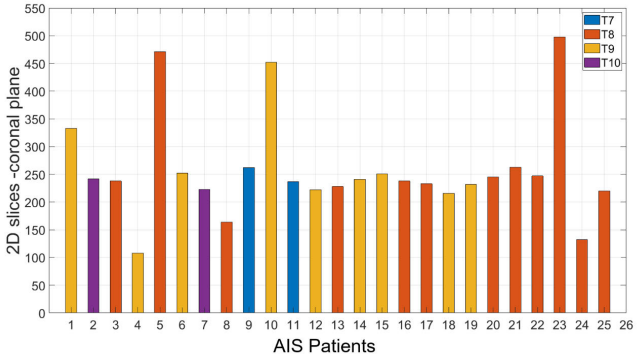

FIGURE 3. Annotated coronal slices per volunteer and vertebral level.

b: AD HOC DATA AUGMENTATION FOR TRAINING

Different types of data augmentation were implemented to include in the training dataset possible image variations specific to the unlabeled vertebral levels in the dataset (i.e. T5, T6, T11, T12). The MRIs were visually inspected and the primary sources of variability among different vertebral levels identified.

The vertebrae morphology captured by the coronal plane varies based on the specific spine levels and the coronal plane intersection angle with the vertebra. The image transformations to compensate for these effects included:

- Rotation and horizontal translation. Each image was rotated and translated by an amount randomly selected between $(-10, 10)$ degrees and $(-10, 10)$ pixels, or between $(-15, 15)$ degrees and $(-15, 15)$ pixels (Figure 4 b).
- Shear deformation. A random shear value between $(0.05, 0.5)$ and between $(-0.05, -0.5)$ was selected for each annotated image (Figure 4 c).
- Coronal plane tilt. Each MRI subvolume was resliced into 2D images 10 times, after tilting the coronal plane at different angles from -15 to 15 degrees, with an increment of 3 degrees each time.

Image intensity changes were also observed, such as varying levels of boundary definition and pixel intensity values surrounding the vertebrae, due to the presence of different tissues adjacent to the vertebral levels. More specifically, the apical vertebrae in the training set were mostly characterised by a black background due to the presence of the lungs, while this was not the case for lower vertebral levels (e.g. T11 and T12) surrounded by soft tissues. To this end, two additional image transformations were implemented (Figure 4 d-e):

- Blurring. A blurring factor randomly chosen between 1 and 3 was applied to each image.
- Background intensity override or background intensity override and blurring. Darker pixels (intensity values lower than 80 on a scale from 0 to 255) at the sides of the annotated vertebra were replaced with a randomly selected grey value between 150 and 170. This solution aimed at forcing the algorithm to generate a solution even if the region to be segmented was not surrounded

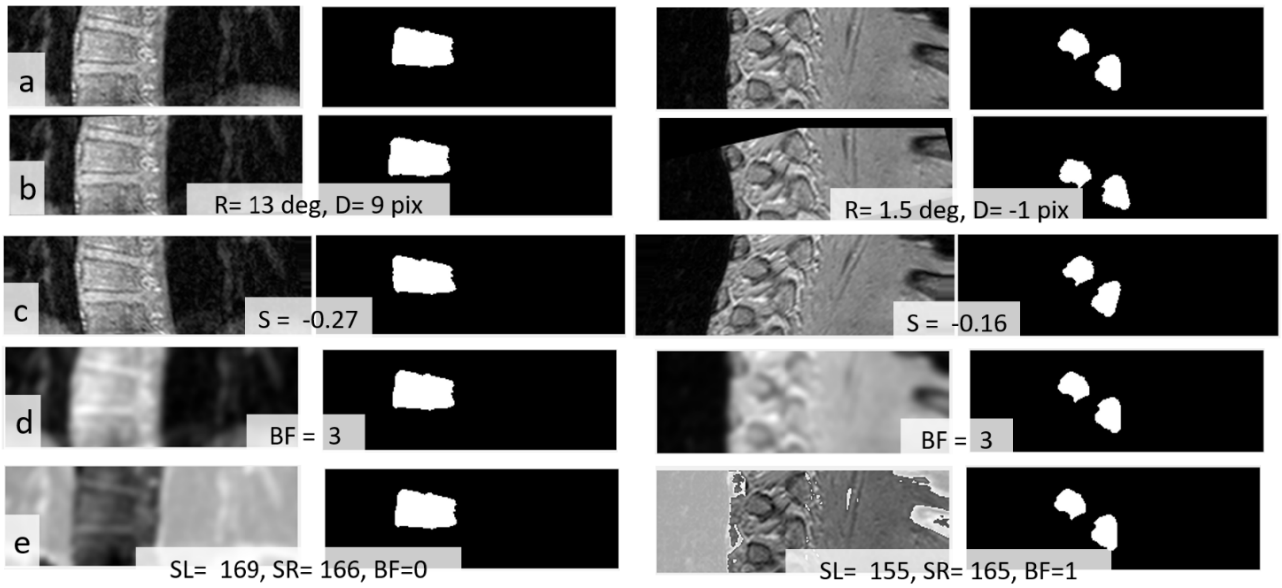

FIGURE 4. Image transformation examples applied to two MRI coronal slices and corresponding labels (a) including the vertebral body and the posterior process (left and right column, respectively). b) Rotation R and horizontal translation D; c) Shear deformation (S=shear factor); d) Blurring (BF = blurring factor); e) Background intensity override (SL/SR = left/right side pixel intensity) and blurring.

by a black background. Subsequently, blurring (with a random factor between 1 and 3) was also applied to some of the images randomly selected.

The ranges of values used for the image transformations described above were selected by visually inspecting the MRI volumes. 90,010 2D images were obtained from the data augmentation process in addition to the original dataset. Black pixel padding was applied to both the original and the newly generated images (and corresponding annotations) to match the largest image in the dataset (118 × 364 pixels, in height and width, respectively). The images (and annotations) were then downscaled, keeping the same image aspect ratio, to 100 × 308 pixels, to reduce the algorithm computation time.

2) CNN ARCHITECTURE AND TRAINING

A UNet architecture [20] was implemented to perform the automatic segmentation of the vertebrae. It included ten blocks of two sequential convolutional layers. After each convolutional layer, batch normalisation [31], ReLu activation and dropout [32] of 10% of the network weights were applied. The optimal algorithm hyperparameters (Table 3) were defined by using the data from 80% of the patients (20 AIS patients) to train the algorithm and the remaining 20% (5 AIS patients) to estimate the algorithm performance. The algorithm was trained using Adam optimiser [33] and the dice loss as cost function [34].

E. CNN TESTING ON AIS PATIENTS AND HEALTHY VOLUNTEERS

1) CNN CROSS-VALIDATION

The algorithm cross-validation enabled partitioning the dataset 5 times into training and test sets, using an 80% to 20% ratio (i.e., 20 to 5 AIS patients, respectively), each time including different AIS patients in the test set. The training and the test sets included on average 46.6 ± 1.67 volumes and 11.4 ± 1.67 volumes, respectively. Data augmentation was applied to all the training volumes as described in Section II-D-1-b.The algorithm was retrained with these new data splits utilising the same network structure and hyperparameters as specified in Table 3, resulting in 5 trained models.

TABLE 3. Optimised algorithm hyperparameters.

Hyperparameters	Value
Learning rate	10^{-3}
Weight decay	10^{-5}
Momentum	0.95
Batch size	8
Epochs	15

2) CNN TESTING WITH SLIDING WINDOW APPROACH FOR VERTEBRAL LEVELS FROM T5-T12

The vertebrae T5-T12 were segmented on the AIS patients and the healthy volunteers using the trained models assigned in the cross-validation and the best performing model in the cross-validation, respectively. Since the MRI volumes included the whole upper body area, the most anterior/posterior slices of the MRI not comprising the spine were removed to reduce computation time. The medial-most and lateral-most MRI slices were then cropped to produce the same fixed boundary size applied to the annotated ground-truth images in Section II-D-1. Black padding was applied at the lateral and medial sides of the images to match the training images' width (i.e. 308 pixels in width).

The next step involved subdividing the MRI coronal slices into patches capturing the vertebrae of interest, as performed in training. To this end, a "sliding window approach" was adopted. A window of fixed dimension (100 pixels) was translated on each coronal slice of the MRI volume with a stride of 10 pixels. Each image part overlapped by the window was resized to 100 × 308 (as the annotated dataset images in Section II-D-1b) and fed to the network. The resulting segmentations were then resized and re-ordered based on their original dimensions and positions on the corresponding MRI, and overlapping pixels (from different segmentations) were summed. The segmentations obtained for each MRI coronal slice were finally combined to form a binary volume and were resized to the original MRI volume voxel dimensions (Figure 5).

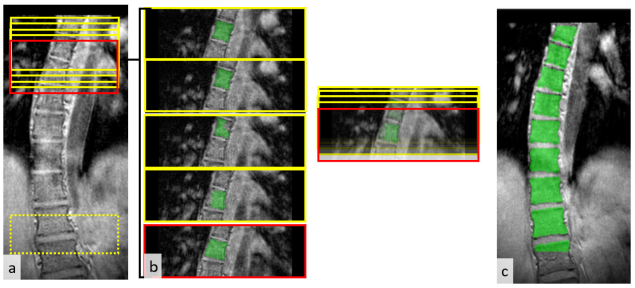

FIGURE 5. Schematic representation of the sliding window approach. In a) a window with fixed dimension and stride is translated through an example MRI coronal slice. In b), all the image regions overlapped by the window are fed to the network and segmented (shown in green). The resulting segmentations are then re-ordered based on their original position and summed. In c) the result after this procedure is applied to the whole MRI slice is shown.

The window dimension selected was an overestimation of the region needed to enclose a vertebra on the coronal plane and combined with a relatively small stride, it ensured that each vertebra (both vertebral body and posterior elements) in the coronal plane could be fully enclosed by the sliding window at least once. The algorithm performance was assessed in relation to the size of the window and stride selected, by testing two additional window-stride combinations: 80 pixels window with 10 pixels stride and 80 pixels window with 2 pixels stride.

3) POST-PROCESSING FOR QUALITATIVE ANALYSIS

Post-processing was implemented for a qualitative analysis of the results. The volumetric segmentation generated as described in Section II-E-2 was smoothed using a standard Gaussian smoothing filter, with the kernel radius parameter (defining the size and strength of the smoothing effect) of 10. Relatively small connected components (below 7000 pixels) were subsequently deleted. These post-processing steps were automatically performed through the "smoothing" function and the "connectivity analysis" module embedded in Imfusion.

Note, however, that the quantitative evaluations reported in Section II-F were performed on the raw results obtained (Section II-E-2), without the post-processing steps here described.

F. CNN AND INTRA-OBSERVER PERFORMANCE EVALUATION

1) EVALUATION METRICS

- The Dice Score Coefficient (DSC) [35] is a quantitative measure of the overlap between two binary masks (Eq 1). It is calculated by superimposing the two masks and computing the number of overlapping pixels multiplied by a factor of 2, divided by the total number of segmented pixels in the two masks. Thus, two perfectly overlapping masks would result in a DSC of 1 (or 100%).

$$\mathrm{DSC} = \frac{2(M_1 \cdot M_2)}{|M_1| + |M_2|} \tag{1}$$

where "·" indicates the dot product, the subscripts refer to the masks 1 and 2, and the absolute value of each mask corresponds to the total number of segmented pixels.

- The Hausdorff distance (HD) [36] measures the maximum distance selected among the minimum distances computed between the boundary pixels of the two binary masks (B_1 and B_2) (Eq 2-4).

$$\mathrm{HD} = \max\left(hd\left(B_1, B_2\right), hd\left(B_2, B_1\right)\right) \tag{2}$$

$$hd\left(B_1, B_2\right) = \max_{b_1 \in B_1} \min_{b_2 \in B_2} \|b_1 - b_2\|^2 \tag{3}$$

$$hd\left(B_2, B_1\right) = \max_{b_2 \in B_2} \min_{b_1 \in B_1} \|b_1 - b_2\|^2 \tag{4}$$

Two additional metrics were developed to further assess the segmentations for each vertebral level and for all the vertebral levels in each MRI.

- The Segmentation Scoring Rate (SSR) is a quantitative metric to compute the rate of correctly segmented 2D coronal slices (Eq 5). For each vertebral level, two separate SSR were computed for the vertebral body (VB) and the posterior process (PP). A score of 0, 0.5 or 1 was assigned by visual inspection whenever the vertebra was not segmented, partially or fully segmented (morphologically clinically acceptable segmentation), respectively. More specifically, to be considered clinically acceptable, the segmentation boundary had to include the cortical bone (Figure 6), with an error tolerance of +/− 1 pixel. Moreover, a 0.5 score was assigned if the vertebra was correctly segmented, but false positives were connected to the prediction; whereas a score of 1 was given when false positives were present but were not connected to the segmentation. The scores assigned for the coronal slices "i" of a vertebral level were then summed and divided by the total number of 2D images where the vertebra part (either VB or PP) was present.

$$\mathrm{SSR} = \frac{\sum_{i=1}^{n} score_i}{n} \tag{5}$$

where n is the number of 2D coronal slices where a specific vertebral level was present.

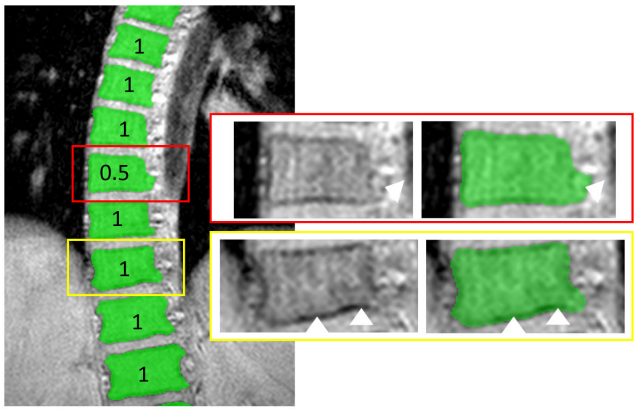

FIGURE 6. Schematic representation of the sliding window approach. In a) a window with fixed dimension and stride is translated through an example MRI coronal slice. In b), all the image regions overlapped by the window are fed to the network and segmented (shown in green). The resulting segmentations are then re-ordered based on their original position and summed. In c) the result after this procedure is applied to the whole MRI slice is shown.

- The Overall Score (OS) is a quantitative measure assessing for each MRI the quality of the segmentations produced for the vertebrae from T5 to T12 (Eq 6- 7). It is computed as the average of the evaluations V_l assigned to each vertebral level in an MRI volume (Eq 7).

$$\mathrm{OS} = \overline{V}_l \tag{6}$$

$$V_l = w_1(SSR_l^{VB}) + w_2\left(SSR_l^{PP}\right) \tag{7}$$

where $\mathrm{SSR}_l^{\mathrm{VB}}$ and $\mathrm{SSR}_l^{\mathrm{PP}}$ are the SSR computed for the VB and the PP of the vertebral level in the MRI, respectively. The weights w_1 and w_2 were set to 0.5. Note that if one is interested only in VB segmentation then $w_1 = 1$ and $w_2 = 0$.

2) CNN PERFORMANCE ON AIS PATIENTS AND HEALTHY VOLUNTEERS

a: CNN CROSS-VALIDATION

The algorithm performance was first evaluated on the vertebral levels available in training. The mean DSC was computed between the 2D segmentations generated by the algorithm and the expert, for each test set and each vertebral level in the test sets (Section II-E-1).

b: SLIDING WINDOW OPTIONS COMPARISON

The different windows-stride dimensions combinations described in Section II-E-2 were tested on 5 MRIs randomly picked from different test sets, ensuring that the combination of MRIs selected had a ground-truth label for each vertebral level. To assess the algorithm performance in relation to the window-stride dimension selected, the mean DSC was computed between the ground-truth segmentation and the corresponding vertebrae in the predicted segmentation.

Since the segmentation obtained from the algorithm comprised the vertebrae from T5 to T12, the vertebral level to be compared with the corresponding (one vertebra) ground-truth was first selected. This selection was performed by extracting all the connected components (along the coronal plane) of the predicted segmentation region enclosed by the bounding box of the corresponding ground-truth annotation plus a small margin (5 pixels in the medial-lateral and superior-inferior directions).

The mean DSC was recomputed for some examples of the annotated vertebrae in the test sets, to account for possible changes in the algorithm performance caused by the sliding window approach.

TABLE 4. Mean DSC for the test set used in each CNN model and for the corresponding vertebral levels. Column 1 details the CNN model. Columns 2-6 report the mean DSC computed over all the images in the corresponding test set and the mean DSC considering only the images of a specific vertebral level (T7, T8, T9 or T10), respectively.

CNN Model	Mean DSC [%]	Mean DSC T7 [%] (images)	Mean DSC T8[%] (images)	Mean DSC T9[%] (images)	Mean DSC T10[%] (images)
1	88.3	-	88.5 (819)	87.5 (448)	89.5 (239)
2	87.9	-	87.8 (941)	88.8 (213)	87.7 (221)
3	89.0	-	89.6 (711)	88.4 (469)	-
4	90.0	90.7 (469)	89.8 (496)	89.0 (463)	-
5	88.0	88.2 (234)	87.4 (159)	88.1 (797)	-

c: CNN TESTING FOR VERTEBRAL LEVELS T5-T12

The 5 MRIs selected in Section II-F-2-b and 3 additional randomly selected MRIs of healthy volunteers were assessed first by comparing the mean DSC and the SSR (Section II-F-1) for the vertebral levels for which a corresponding ground-truth was available. As described in Section II-F-2-b, to enable this evaluation, the vertebral level corresponding to the ground-truth vertebra was extracted from the predicted segmentation, which included all vertebral levels T5-T12.

The dataset was then comprehensively evaluated computing the SSR for each vertebral level (with or without comparison of corresponding ground-truth label) and the OS to obtain the overall performance of each MRI analysed.

3) INTRA-OBSERVER AND INTER-OBSERVER VARIABILITY TESTS

The intra-observer and the inter-observer variability were assessed to compare the experts and the algorithm performance and to understand the maximum achievable performance for the algorithm. The first expert (H.J.) contoured the apical vertebra of one of the AIS patient three times, at different time points. The DSC was computed between the 2D segmentations of the first and the second contoured

TABLE 5. Mean DSC for the test set used in each CNN model and for the corresponding vertebral levels. Columns 1-3 detail the CNN model, the volunteer ID and the vertebral level, respectively. Columns 4-7 report the mean DSC obtained for the three different Window-Stride combinations and when the sliding window approach was not applied. The cases where the Window-Stride combinations obtained a significant mean DSC difference are highlighted.

			Mean DSC[%]			
CNN Model	Volunteer ID	Vertebra	Window - Stride [100-10]	Window-Stride [80-10]	Window-Stride [80-2]	No Sliding Window
1	5	T6	87.1	87.1	87.0	-
	5	T10	90.5	90.5	90.5	89.5
	5	T11	90.4	90.4	90.4	-
2	6	T6	83.8	84.1	84.7	-
	6	T8	89.0	89.0	88.9	88.5
	6	**T12**	**85.3**	**84.5**	**86.0**	-
3	11	T9	92.5	91.9	92.0	90.9
4	14	T7	91.5	91.5	91.2	91
5	**3**	**T5**	**82.4**	**81.8**	**83.0**	-
	3	**T6**	**77.1**	**77.2**	**84.9**	-
	3	T8	87.3	87.2	87.8	83.6
	3	T11	85.6	84.7	85.3	-

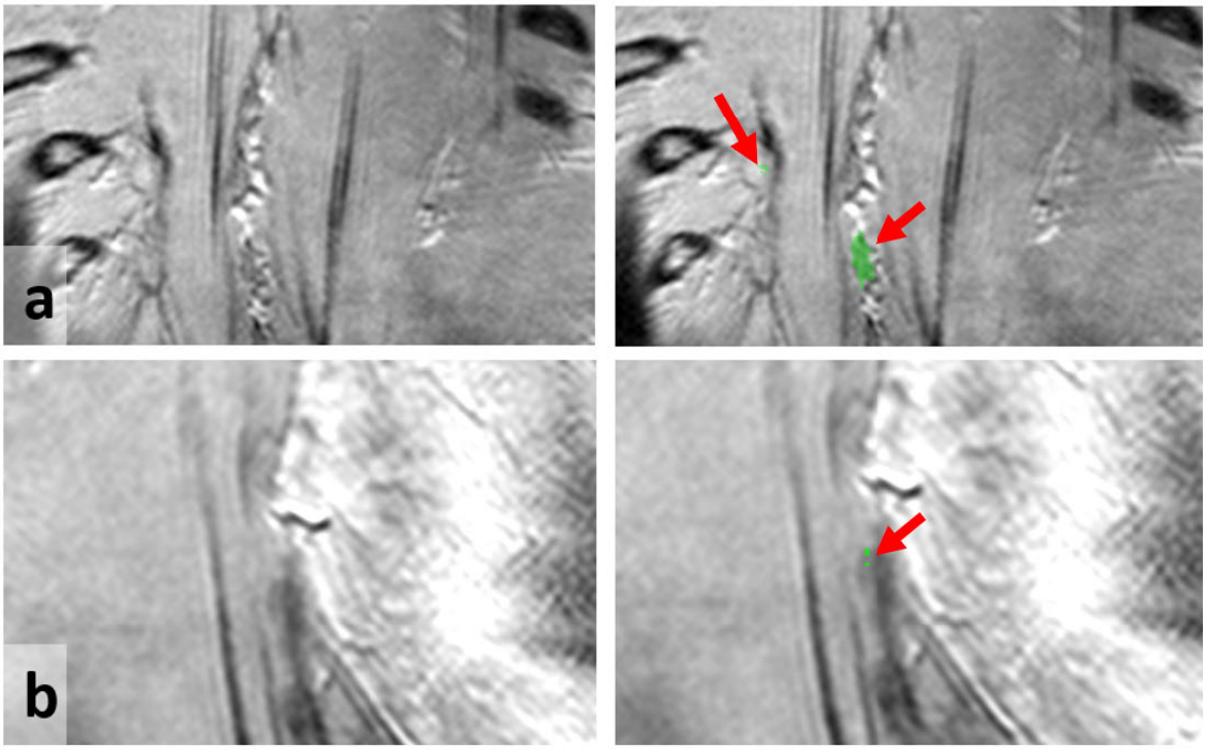

FIGURE 7. The most posterior coronal slice assessed through the SSR metric (a) and the actual most posterior slice contoured by the expert (b) (The actual posterior-most coronal slice was 12 MR images posterior to the one assessed with SSR). In the left and right column, the MRI and the corresponding superimposed ground-truth segmentations (in green and highlighted by arrows) are shown, respectively.

volumes, the second and the third contoured volumes, and the first and the third contoured volumes.

The second expert (see Section II-C) outlined the same vertebra three times, at different time points and the DSC was computed as for the first expert. The same procedure was followed to assess the intra-observer variability for more superior and inferior verbal levels of the torso (T6 and T11).

Finally, the inter-observer variability was computed between the two experts for the apical vertebra, comparing the labels generated during three contouring sessions for one

TABLE 6. Comparison between the DSC, SSR and HD evaluation computed for each volunteer individually. Columns 1-5 detail the volunteer type (AIS or healthy), the CNN model, the volunteer ID, the vertebral level, and the number of pixels contained in the corresponding ground-truth (GT), respectively. Columns 6-8 report the mean DSC considering the vertebral body (VB) and the posterior process (PP) individually and combined (TOT), excluding from the evaluation the most anterior-posterior MRI slices with not well-defined vertebra boundaries. Columns 10-15 report the corresponding SSR and mean HD obtained. Column 9 details the mean DSC with superscript* calculated (for the vertebral body and posterior process) over all the coronal slices where the vertebrae were present, even if not well defined. The cases corresponding to the lowest SSR performance (over the whole vertebra) are highlighted.

					Window-Stride [100-10]									
	#CNN Model	Volunteer ID	Vertebra	# Pixel GT	Mean DSC^{VB} [%]	Mean DSC^{PP} [%]	Mean DSC^{T} ^{OT} [%]	Mean DSC^{T} ^{OT*}[%]	SSR^{VB} [%]	SSR^{PP} [%]	SSR^{TO} ^{T}[%]	Mean HD^{VB}[mm]	Mean HD^{PP} [mm]	Mean HD^{TOT} [mm]
AIS	1	5	T6	121040	93.9	88.3	91.1	87.2	96.5	96.6	96.0	1.98± 1.24	2.85± 1.69	2.49± 1.57
		5	T10	164903	94.8	90.9	92.9	90.6	100	99.3	99.6	2.20± 0.91	2.51± 1.84	2.39± 1.53
		5	T11	185878	95.7	90	92.9	90.4	97.7	92.6	95.2	2.31± 1.62	2.93± 1.80	2.67± 1.75
	2	6	T6	132428	92.8	83.8	88.3	83.8	100	99.2	99.6	2.98± 0.92	3.38± 1.14	3.21± 1.06
		6	T8	141728	94.0	88.6	91.3	89	94.7	100	97.4	1.99± 0.87	3.00± 1.67	2.55± 1.46
		6	**T12**	**206366**	**95.2**	**82.4**	**88.8**	**85.3**	**100**	**74.3**	**87.2**	**2.30± 0.77**	**4.59± 5.26**	**3.65± 4.20**
	3	11	T9	158271	93.3	93.3	93.3	92.5	99.1	100	99.6	1.57± 1.88	1.90± 1.21	1.75± 1.55
	4	14	T7	163931	95.1	93.3	94.2	91.5	100	97.9	99.0	1.36± 1.00	1.72± 0.89	1.57± 0.95
	5	3	T5	83603	93.4	80.6	87.0	82.5	97.5	90.9	94.2	2.15± 0.81	3.67± 2.61	3.07± 2.22
		3	T6	107420	92.8	86.8	89.8	77.2	97.6	99.2	98.4	2.22± 1.39	3.47± 2.37	2.94± 2.10
		3	T8	127009	92.9	89.5	91.2	87.4	100	100	100	1.98± 1.56	3.07± 1.22	2.59± 1.48
		3	**T11**	**164625**	**93.2**	**84.1**	**88.7**	**85.7**	**95.7**	**71.5**	**83.6**	**3.38 ±1.21**	**5.34± 3.82**	**4.46± 3.09**
Healthy	4	1	T9	100243	92.2	93	92.6	91.2	100	96.7	98.3	1.46± 1.19	2.06± 1.65	1.81± 1.49
		5	T9	116564	92.7	93.8	93.2	92.4	100	100	100	1.58± 1.11	1.83± 1.18	1.72± 1.15
		10	T9	127199	93.1	93.1	93.1	92.5	100	94.5	97.3	1.94± 1.35	1.57± 1.06	1.78± 1.22

volume and the labels outlined during a single contouring session for two additional volumes.

III. RESULTS

A. CNN CROSS-VALIDATION

The 5 CNN models showed similar performance, with mean DSC between 88% and 90% (Table 4). The same trend was observed analysing each vertebral level individually, despite the vertebral levels T7 and T10 were underrepresented in the training set.

B. SLIDING WINDOW OPTIONS COMPARISON

The results obtained for different window-stride combinations did not show a significant variation in most of the cases analysed (Table 5) (the mean DSC variation was typically less than 1%). The window-stride combination [80-2] showed a performance increase of about 1-2% in a few cases and of about 8% in one case (highlighted in Table 5). The overall computation time (including opening the dicom file and writing the segmentation generated) was about 2 minutes for the window-stride combinations [100-10] and [80-10], and 12 minutes when a stride of 2 was utilised. All computations were performed on an online cloud (RONIN, New Jersey, United States) using 2 GPUs NVIDIA Tesla M60 (NVIDIA, Santa Clara, California, United States).

C. CNN TESTING FOR VERTEBRAL LEVELS T5-T12

Table 6 compares the DSC, the SSR and the HD evaluation. The metrics were computed for the vertebral body and the posterior process, considered as individual parts

and combined. The MRI coronal slices for which the segmentation was assessed were selected by visual inspection, discarding the most anterior and posterior coronal slices where the boundaries of a given vertebra could not be confidently identified (Figure 7). The mean DSC computed over all the coronal slices where the vertebrae were present (even if not well defined) is also reported in Table 6 with superscript *.

For AIS volunteers, the SSR values obtained were between 83% and 100% (where 100% implied that all the segmented slices considered were clinically acceptable), with the lowest values corresponding to the vertebral levels T11 and T12 (highlighted in Table 6). This trend was confirmed by the HD, but not by the mean DSC obtained. Since the vertebrae T11 and T12 were the largest vertebral levels in the dataset (about double the size of a T5 or a T6), segmentation errors would be penalised less compared to the other vertebral levels.

The SSR reported for each vertebral level (over 5 and 3 MRIs for AIS and healthy participants, respectively) confirmed that a significantly lower performance was obtained for the levels T11 and T12, especially for the posterior process (Table 7). Compared to AIS patients, for healthy participants, the mean SSR was slightly lower for the vertebrae T7-T10, but it significantly improved for the vertebrae T5, T11 and T12 (by 4.8, 7.8 and 12.5 %, respectively). When all the vertebral levels were considered, healthy participants showed a higher performance compared to AIS patients (Table 8).

TABLE 7. Mean overall SSR reported for each vertebral level computed over 5 and 3 MRIs of AIS and healthy participants, respectively. Columns 1-2 describe the volunteer type (AIS or healthy), respectively. Columns 3-5 report the mean SSR computed for the vertebral bodies (VB), the posterior process (PP) or the two combined (TOT) respectively, for the corresponding vertebral level. The cases corresponding to the lowest SSR performance (over the whole vertebra) are highlighted.

	Vertebral level	Mean SSR^{VB} [%]	Mean SSR^{PP} [%]	Mean SSR^{TOT} [%]
AIS	**T5**	**90.7**	**95.0**	**92.8**
	T6	98.1	98.4	98.3
	T7	99.3	99.2	99.2
	T8	98.9	99.5	99.2
	T9	99.0	99.9	99.4
	T10	97.1	97.8	97.4
	T11	**88.9**	**83.9**	**86.4**
	T12	**77.2**	**70.5**	**73.8**
Healthy	T5	98.3	96.8	97.6
	T6	97.9	96.5	97.2
	T7	99.2	98.6	98.9
	T8	99.6	97.9	98.7
	T9	100	97.0	98.5
	T10	99.0	94.5	96.7
	T11	**98.7**	**89.7**	**94.2**
	T12	**94.0**	**78.6**	**86.3**

TABLE 8. OS values for AIS and healthy participants.

	Volunteer ID	OS[%]
AIS	5	96.1
	6	94.1
	11	94.1
	14	88.5
	3	92.3
Healthy	1	96.5
	5	96.2
	10	99.0

D. INTRA-OBSERVER AND INTER-OBSERVER VARIABILITY TESTS

The intra/inter-observer mean DSC was on average 86.7% ± 2.8% and 86.2% ± 2.6%, respectively. Table 9 describes these results for all the cases analysed. The higher DSC of the experts was typically associated with the consistency in selecting the first and the last MRI images that were considered contourable within a volume.

E. QUALITATIVE ANALYSIS AFTER POST-PROCESSING

Figures 8 and 9 show examples of raw and the post-processed segmentations (Section II-E-3) of AIS patients and healthy volunteers. The post-processing resulted in deleting the false positives generated in the raw solution and enabled connecting incomplete adjacent segmentations. However, the smoothing algorithm utilised also caused an undesired merging of the boundaries of adjacent vertebrae, due to the small distance between them. Furthermore, filtering small components resulted in deleting segmentation parts of the vertebrae.

IV. DISCUSSION

In this paper, we present a deep-learning technique to reconstruct the thoracic spine levels T5-T12, using the ground-truth label data of a single vertebra per patient. Without loss of generality, the sliding window approach and the data augmentation utilised here could be applied to any deep learning algorithm for spine segmentation. The implemented data augmentation consisted of commonly used image processing transformations, and the selected parameters were purely derived by visual inspection. The sliding window size was arbitrarily chosen, solely considering that it had to exceed the dimension of the individual vertebrae along the coronal plane and, the algorithm performance was proved to be robust to significant changes in the sliding window dimension. The stride dimension most strongly influenced the segmentation processing time since it defined the number of image patches extracted from a given MRI volume. Utilising a smaller stride (e.g. 2 pixels) was time expensive, but it could lead to higher segmentation performance for the vertebral levels not represented in the training dataset. It ensured in fact that the network could be fed with multiple patches capturing the vertebra, or different regions of it, at varying locations within the patch. At the current stage, the stride can be selected depending on the specific requirements of the segmentation

TABLE 9. Mean intra/inter-observer mean DSC computed for two experts. Columns 1-2 describe the volunteer ID and the vertebral level, respectively. Column 3-5 report the intra/inter-observer mean DSC computed for two experts.

Volunteer ID	Vertebra	Intra-observer A Mean DSC[%]			Intra-observer B Mean DSC[%]			Inter-observer A-B Mean DSC[%]								
		1^A - 2^A	2^A - 3^A	3^A - 1^A	1^B – 2^B	1^A –2^B	2^A – 2^B	1^A – $!^B$	2^A – 1^B	3^A– 1^B	1^A – 2^B	2^A – 2^B	3^A – 2^B	1^A – 3^B	2^A – 3^B	3^A– 3^B
3	T8	86.3	87.9	84.2	81.4	89.8	81.6	85.7	87.8	86.0	79.8	82.2	81.2	80.5	83.3	82.9
5	T10	-	-	-	-	-	-	87.4	-	-	-	-	-	-	-	-
6	T8	-	-	-	-	-	-	88.0	-	-	-	-	-	-	-	-
3	T6	-	-	-	83.3	87.8	89.5	-	-	-	-	-	-	-	-	-
3	T11	-	-	-	88.1	91.7	89.0	-	-	-	-	-	-	-	-	-

The superscripts refer to observers A and B and the CNN C; 1, 2, 3 indicate the segmentation produced during the first, second or third contouring session, respectively.

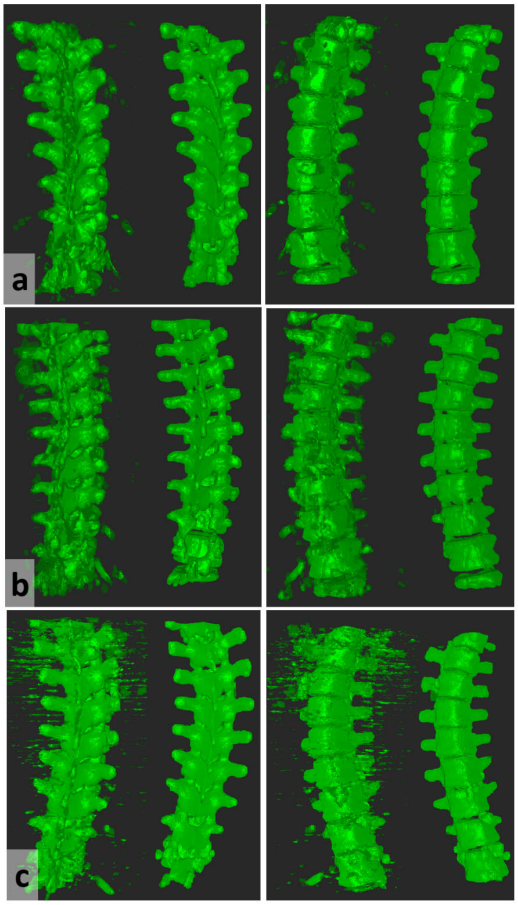

FIGURE 8. 3D reconstructions of AIS patients vertebral levels T5-T12 before and after post-processing (leftmost and rightmost sides, respectively). Figures a, b, c correspond to 3 AIS patients. The left/right columns of images correspond to the posterior/anterior views of the reconstructions.

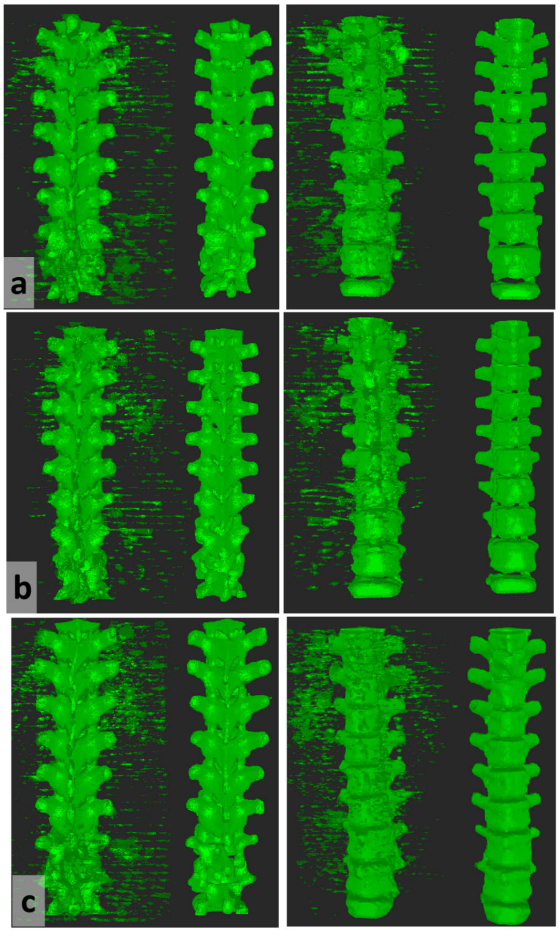

FIGURE 9. 3D reconstructions of healthy volunteers vertebral levels T5-T12 before and after post-processing (leftmost and rightmost sides, respectively). Figures a, b, c correspond to 3 healthy volunteers. The left/right columns of images correspond to the posterior/anterior views of the reconstructions.

application to find the best trade-off between processing time and segmentation accuracy. In the future, this limitation could be overcome through the use of additional training data.

Furthermore, since the algorithm was typically conservative in the vertebral anatomy detection, in this work, the final segmentation generated through the sliding window approach was defined as the union of all the segmentations generated on the image patches. The effectiveness of this choice was also proved through the fact that using a small stride, and thus

considering solutions of an increased number of overlapping image patches, led to enhanced algorithm performance.

The segmentation approach used in this work was successful for both healthy volunteers and AIS patients. The reconstructions of healthy spines were, on average, more accurate, even though the algorithm was trained uniquely on AIS patient's data. This result can be explained by the patient-specific spinal deformity and/or intra-vertebral rotations in AIS anatomy that created enhanced variability in the vertebral appearance, to which the algorithm was not always robust.

The 3D reconstruction of the vertebral anatomy here presented can be used to simulate patient-specific mechanics in biomechanical investigations or to examine the AIS pathology and, eventually, volumetrically analyse AIS vertebral growth and deformity progression by comparing the vertebral anatomy reconstructions in the sequential MRIs of AIS and healthy volunteers. This study is particularly relevant since the only alternative to reconstruct the vertebral anatomy during the AIS disease progression is using EOS biplanar X-ray imaging [37] (EOS Imaging, Paris, France) that, being based on generic shape fitting algorithms, cannot generate a morphologically accurate model of the spine. The main requirement defining the usability of the segmentations generated to analyse the pathology is the ability of accurately reconstructing the anatomy of the vertebral bodies, which are the elements contributing the most to the spine deformity since they are the largest vertebral structures. The SSR metric introduced in this paper indicated that clinically acceptable morphological reconstructions were typically generated for both the vertebral body and posterior process, with some exceptions at the vertebral levels T11 and T12. This effect was caused by the substantial variation in size and morphology of these spinal levels, in particular for the posterior process, when compared to other thoracic vertebrae. In the future, to enhance the algorithm performance, additional vertebral level(s) representative of the lower spine should be introduced in the training set, accounting thus for their shape difference.

In addition to the SSR, the segmentation results were also obtained using the standard DSC. The results achieved by the two metrics were not always directly proportional, since the DSC evaluates the segmentations at a pixel-level, whereas the SSR at a structural/morphological-level. In particular, the DSC failed to identify the lower segmentation quality of T11 and T12. Segmentation errors resulted in a smaller penalty for these two vertebrae, due to their significantly larger size compared to other vertebral levels.

The mean DSC obtained for AIS and healthy volunteers' vertebral bodies (about 93-94% DSC) is comparable to the current state-of-the-art [27] and significantly higher relative to the previous study focusing on AIS patients [25], which reported DSC values between 80 and 90%. A performance similar to this paper was also reported by Korez *et al.* [28] and by Lu *et al.* [26]. However, one should note that this comparison is purely qualitative and should not be taken literally since different test sets were utilised to generate the results reported by the different works. These results depend not only on the effectiveness of the segmentation method applied but also on the specific MRI protocol employed, the vertebral levels segmented, the volunteer group considered and the quality of the ground-truth annotations provided.

In this paper, the quantitative segmentation results were computed on the ''raw'' algorithm segmentations (without post-processing). This choice was made to evaluate the algorithm performance avoiding a possible bias specific to the post-processing steps selected to smooth and remove reconstruction artefacts. We also applied a simple, automatic post-processing procedure to several segmentations for comparison purposes. This process generated a smooth and well-defined 3D reconstruction of the anatomy, but it also led to sub-optimal results in terms of segmentation quality. For these reasons, in future work, the segmentation post-processing should be refined for this specific task.

The primary source of false positives in the ''raw'' segmentation was caused by the fact that many patches which did not contain any vertebral anatomy were fed to the algorithm. Differently from training and cross-validation, at test-time, the sliding window approach was applied to the whole ROI containing the vertebral levels T5-T12, regardless of the presence of the vertebrae in the image patches generated. In the future, the patches containing the vertebrae to be segmented could be automatically selected using a trained classifier [38], [39] before the segmentation algorithm is applied.

V. CONCLUSION

In this paper, we propose a data-efficient method to automatically segment the thoracic spine (T5-T12) using a state-of-the-art CNN trained with a labelled dataset limited to the AIS patients' apical vertebrae. Furthermore, it is the first MRI-based method reconstructing the whole vertebra bony anatomy, including the vertebral body and the more complex posterior process. The implemented method showed a performance comparable to 2 experts, achieving a mean dice score coefficient of 87%±4.3 for the AIS patient vertebral levels T5-T12. This approach was also robust for auto segmentation of healthy volunteers data. On average, 93% and 97% of the MRI segmented slices were considered clinically acceptable morphological reconstructions of AIS and healthy volunteer vertebrae, respectively. The proposed algorithm can be considered as the first step towards MRI-based screening for AIS.

REFERENCES

[1] M. R. Konieczny, H. Senyurt, and R. Krauspe, ''Epidemiology of adolescent idiopathic scoliosis,'' *J. Children's Orthopaedics*, vol. 7, no. 1, pp. 3–9, 2013.
[2] J. P. Horne, R. Flannery, and S. Usman, ''Adolescent idiopathic scoliosis: Diagnosis and management,'' *Amer. Family Physician*, vol. 89, no. 3, pp. 193–198, 2014.
[3] C. M. Ronckers, C. E. Land, J. S. Miller, M. Stovall, J. E. Lonstein, and M. M. Doody, ''Cancer mortality among women frequently exposed to radiographic examinations for spinal disorders,'' *Radiat. Res.*, vol. 174, no. 1, pp. 83–90, Jul. 2010.

[4] M. R. Schmid and J. Hodler, *Imaging Studies*. Berlin, Germany: Springer, 2008.
[5] D. J. Brenner and E. J. Hall, "Computed tomography—An increasing source of radiation exposure," *New England J. Med.*, vol. 357, no. 22, pp. 2277–2284, 2007.
[6] A. N. Larson, B. A. Schueler, and J. Dubousset, "Radiation in spine deformity: State-of-the-art reviews," *Spine Deformity*, vol. 7, no. 3, pp. 386–394, May 2019.
[7] *Health Risks From Exposure to Low Levels of Ionizing Radiation: BEIR VII Phase 2*, Nat. Acad. Sci., Washington, DC, USA, 2006.
[8] C. L. Hoad, A. L. Martel, R. Kerslake, and M. Grevitt, "A 3D MRI sequence for computer assisted surgery of the lumbar spine," *Phys. Med. Biol.*, vol. 46, no. 8, pp. N213–N220, Aug. 2001.
[9] D. Zukić, A. Vlasák, J. Egger, D. Hořínek, C. Nimsky, and A. Kolb, "Robust detection and segmentation for diagnosis of vertebral diseases using routine MR images," *Comput. Graph. Forum*, vol. 33, no. 6, pp. 190–204, Sep. 2014.
[10] M. Bronstein, J. Favre, K. Hormann, M. Rak, K. Engel, and K. D. Tönnies, "Closed-form hierarchical finite element models for part-based object detection," in *Proc. Int. Symp. Vis., Model. Vis.*, 2013, pp. 137–144.
[11] A. Neubert, J. Fripp, C. Engstrom, R. Schwarz, L. Lauer, O. Salvado, and S. Crozier, "Automated detection, 3D segmentation and analysis of high resolution spine MR images using statistical shape models," *Phys. Med. Biol.*, vol. 57, no. 24, pp. 8357–8376, Dec. 2012.
[12] S. Kadoury, H. Labelle, and N. Paragios, "Spine segmentation in medical images using manifold embeddings and higher-order MRFs," *IEEE Trans. Med. Imag.*, vol. 32, no. 7, pp. 1227–1238, Jul. 2013.
[13] M. S. Aslan, A. Shalaby, and A. A. Farag, "Clinically desired segmentation method for vertebral bodies," in *Proc. IEEE 10th Int. Symp. Biomed. Imag.*, Apr. 2013, pp. 840–843.
[14] M. S. Aslan, A. Shalaby, H. Abdelmunim, and A. A. Farag, "Probabilistic shape-based segmentation method using level sets," *IET Comput. Vis.*, vol. 8, no. 3, pp. 182–194, Jun. 2014.
[15] C. Chu, D. L. Belavý, G. Armbrecht, M. Bansmann, D. Felsenberg, and G. Zheng, "Fully automatic localization and segmentation of 3D vertebral bodies from CT/MR images via a learning-based method," *PLoS ONE*, vol. 10, no. 11, Nov. 2015, Art. no. e0143327.
[16] B. Gaonkar, Y. Xia, D. S. Villaroman, A. Ko, M. Attiah, J. S. Beckett, and L. Macyszyn, "Multi-parameter ensemble learning for automated vertebral body segmentation in heterogeneously acquired clinical MR images," *IEEE J. Transl. Eng. Health Med.*, vol. 5, pp. 1–12, 2017.
[17] F. Fallah, S. S. Walter, F. Bamberg, and B. Yang, "Simultaneous volumetric segmentation of vertebral bodies and intervertebral discs on fat-water MR images," *IEEE J. Biomed. Health Informat.*, vol. 23, no. 4, pp. 1692–1701, Jul. 2019.
[18] D. Gaweł, P. Główka, T. Kotwicki, and M. Nowak, "Automatic spine tissue segmentation from MRI data based on cascade of boosted classifiers and active appearance model," *BioMed Res. Int.*, vol. 2018, pp. 1–13, Apr. 2018.
[19] M. Vania, D. Mureja, and D. Lee, "Automatic spine segmentation from CT images using convolutional neural network via redundant generation of class labels," *J. Comput. Des. Eng.*, vol. 6, no. 2, pp. 224–232, Apr. 2019.
[20] O. Ronneberger, P. Fischer, and T. Brox, "U-Net: Convolutional networks for biomedical image segmentation," pp. 1–8, 2015, *arXiv:1505.04597*. [Online]. Available: https://arxiv.org/abs/1505.04597
[21] Y. J. Kim, B. Ganbold, and K. G. Kim, "Web-based spine segmentation using deep learning in computed tomography images," *Healthcare Informat. Res.*, vol. 26, no. 1, p. 61, 2020.
[22] N. Lessmann, B. van Ginneken, P. A. de Jong, and I. Išgum, "Iterative fully convolutional neural networks for automatic vertebra segmentation and identification," *Med. Image Anal.*, vol. 53, pp. 142–155, Apr. 2019.
[23] D. Shi, Y. Pan, C. Liu, Y. Wang, D. Cui, and Y. Lu, "Automatic localization and segmentation of vertebral bodies in 3D CT volumes with deep learning," in *Proc. ACM Int. Conf. Ser.*, 2018, pp. 42–46.
[24] N. Masuzawa, Y. Kitamura, K. Nakamura, S. Iizuka, and E. Simo-Serra, "Automatic segmentation, localization, and identification of vertebrae in 3D CT images using cascaded convolutional neural networks," 2020, *arXiv:2009.13798*. [Online]. Available: http://arxiv.org/abs/2009.13798
[25] N. Guerroumi, C. Playout, C. Laporte, and F. Cheriet, "Automatic segmentation of the scoliotic spine from mr images," in *Proc. IEEE 16th Int. Symp. Biomed. Imag. (ISBI)*, Apr. 2019, pp. 480–484.
[26] J.-T. Lu, S. Pedemonte, B. Bizzo, S. Doyle, K. P. Andriole, M. H. Michalski, R. G. Gonzalez, and S. R. Pomerantz, "Deep spine: Automated lumbar vertebral segmentation, disc-level designation, and spinal stenosis grading using deep learning," in *Proc. 3rd Mach. Learn. Healthc. Conf.*, Jul. 2018, pp. 403–419.
[27] M. Rak, J. Steffen, A. Meyer, C. Hansen, and K. Tönnies, "Combining convolutional neural networks and star convex cuts for fast whole spine vertebra segmentation in MRI," *Comput. Methods Programs Biomed.*, vol. 177, pp. 47–56, Aug. 2019.
[28] R. Korez, B. Likar, F. Pernuš, and T. Vrtovec, *Model-Based Segmentation of Vertebral Bodies From MR Images With 3D CNNS* (Lecture Notes in Computer Science: Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics). Cham, Switzerland: Springer, 2016.
[29] H. Anitha and G. K. Prabhu, "Identification of apical vertebra for grading of idiopathic scoliosis using image processing," *J. Digit. Imag.*, vol. 25, no. 1, pp. 155–161, Feb. 2012.
[30] H. A. Jennings, C. A. Grant, M. J. Pearcy, R. D. Labrom, G. N. Askin, and J. P. Little, "An MRI-based technique to assess progressive changes in 3D vertebral morphometry in adolescent idiopathic scoliosis," in *Proc. 30th Annu. Sci. Meeting Spine Soc. Aust.*, Apr. 2019.
[31] S. Ioffe and C. Szegedy, "Batch normalization: Accelerating deep network training by reducing internal covariate shift," 2015, *arXiv:1502.03167*. [Online]. Available: http://arxiv.org/abs/1502.03167
[32] N. Srivastava, G. Hinton, A. Krizhevsky, I. Sutskever, and R. Salakhutdinov, "Dropout: A simple way to prevent neural networks from overfitting," *J. Mach. Learn. Res.*, vol. 15, no. 1, pp. 1929–1958, 2004.
[33] D. P. Kingma and J. L. Ba, "Adam: A method for stochastic gradient descent," in *Proc. ICLR Int. Conf. Learn. Represent.*, 2015, pp. 1–5.
[34] F. Milletari, N. Navab, and S.-A. Ahmadi, "V-net: Fully convolutional neural networks for volumetric medical image segmentation," in *Proc. 4th Int. Conf. 3D Vis. (3DV)*, Oct. 2016, pp. 565–571.
[35] L. R. Dice, "Measures of the amount of ecologic association between species," *Ecology*, vol. 26, no. 3, pp. 297–302, 1945.
[36] R. T. Rockafellar and R. J.-B. Wets, "Variational analysis (book)," *J. Phys. Chem. A*, vol. 37, no. 1, p. 117, 1998.
[37] E. Melhem, A. Assi, R. El Rachkidi, and I. Ghanem, "EOS biplanar X-ray imaging: Concept, developments, benefits, and limitations," *J. Children's Orthopaedics*, vol. 10, no. 1, pp. 1–14, Feb. 2016.
[38] Y. Zhou, Y. Liu, Q. Chen, G. Gu, and X. Sui, "Automatic lumbar MRI detection and identification based on deep learning," *J. Digit. Imag.*, vol. 32, no. 3, pp. 513–520, 2019.
[39] A. Jamaludin, M. Lootus, T. Kadir, A. Zisserman, J. Urban, M. Battié, I. McCall, and G. Consortium, "Automation of reading of radiological features from magnetic resonance images (MRIs) of the lumbar spine without human intervention is comparable with an expert radiologist," *Eur. Spine J.*, vol. 32, no. 3, pp. 513–520, 2017.