Automatic Localization of the Lumbar Vertebral Landmarks in CT Images with Context Features

Dimitrios Damopoulos[1], Ben Glocker[2], Guoyan Zheng[1]

[1] Institute for Surgical Technology and Biomechanics, University of Bern, Bern, Switzerland
dimitrios.damopoulos@istb.unibe.ch
[2] Biomedical Image Analysis Group, Imperial College London, London, United Kingdom

Abstract. A recent research direction for the localization of anatomical landmarks with learning-based methods is to explore ways to enrich the trained models with context information. Lately, the addition of context features in regression-based approaches has been tried in the literature. In this work, a method is presented for the addition of context features in a regression setting where the locations of many vertebral landmarks are regressed all at once. As this method relies on the knowledge of the centers of the vertebral bodies (VBs), an automatic, endplate-based approach for the localization of the VB centers is also presented.
The proposed methods are evaluated on a dataset of 28 lumbar-focused CT images. The VB localization method detects all of the lumbar VBs of the testing set with a mean localization error of 3.2 mm. The multi-landmark localization method is tested on the task of localizing the tips of all the inferior articular processes of the lumbar vertebrae, in addition to their VB centers. The proposed method detects these landmarks with a mean localization error of 3.0 mm.

Keywords: Regression · Localization ·Lumbar · Vertebral body · Inferior articular process.

1 Introduction

Back pain in general and low back pain in particular constitutes a major public health problem, exhibiting epidemic proportions [1]. The computer-assisted diagnosis of pathologies of the lumbar spine involves the analysis of images coming from a series of standard imaging modalities. Computed tomography (CT) images can be used for the diagnosis of spondylolysis, spondylolisthesis and osteoporosis, as this imaging modality permits the measurement of the bone mineral density of the vertebral bodies (VBs). This work focuses on the task of the localization of the lumbar VBs in CT images and the localization of key landmarks on the vertebral processes.

The proposed framework can facilitate subsequent automated procedures, such as the segmentation of the vertebrae, the automatic assessment of skeletal vertebral pathologies and the analysis of the spinal shape. In the case of vertebral segmentation, a large number of proposed methods employ some form of Active Shape Models (ASMs) or Active Appearance Model (AAMs) ([1,2]). The initialization step of these model-

based approaches is typically based on the localization of the centers of the VBs. Using more landmarks than just the center of the VBs can add robustness to this initialization step. Furthermore, the detection of vertebral landmarks can function as the building block for the automated assessment of pathologies concerning the global spinal shape (scoliosis, lordosis) and the grading of spondylolisthesis: In [4], an automated method for the measurement of spondylolisthesis was presented, based on the identification of the endplate regions. For this application, the localization of the edges of the endplates in the coronal direction could provide a more direct method for the measurement of the anterior shift.

Localization of anatomical landmarks is a fundamental problem in medical image analysis and a plethora of methods have been proposed in the literature. In recent years, these tend to be based on machine learning tools and they can be roughly categorized into classification-based methods and regression-based methods. A popular research direction is the addition of context information in the model that is constructed by these learning-based methods. For the problem of object segmentation, a principled method for achieving so is the Auto-context framework ([5]). Recently, there have been attempts to apply this framework for the localization of landmarks with random forest regressors. In particular, in [6] the authors showed that the extraction of context features from the distance maps of a traditional random forest regressor can improve the landmark localization accuracy. Following this research direction, in the present work this method of adding context information is applied to a multiple-landmark localization task, where the locations of more than one landmark are regressed all at once by random forest regressors. We show that the proposed method is able to detect robustly key landmarks of the vertebrae, despite the similar appearance of neighboring vertebra. As the proposed method assumes that the centers of the VBs have been already detected, we also present an endplate-based method for the detection of the lumbar VB centers. We evaluate both of the methods in a dataset of 28 lumbar-focused CT images.

2 Method

The proposed framework consists of two modules. The first module deals with the localization of the VBs and the estimation of the pose of the vertebrae. It performs this task via the detection of the vertebral endplates on a spline-based unwrapping of the input image. The second module deals with the localization of key landmarks of the vertebrae, based on the estimation of the VB centers and the vertebral pose by the first module. It employs two levels of random forest regressors. The two modules are described in sections 2.1 and 2.2. For the rest of this section, it is assumed that a number of CT images of the lumbar spine are available for training. A training image will be referred to using the notation:

$$I_i: \Omega_i \subset \mathbb{R}^3 \to \mathbb{R}, i \in \{1, \cdots, N\} \quad (1)$$

A testing image will be denoted with I_T. Every training image is accompanied with annotations of the centers of all the VBs within the field of view. We denote the set of the annotations of the training image I_i with:

$$A_i = \{(\boldsymbol{c}_1, v_1), \cdots, (\boldsymbol{c}_{m_i}, v_{m_i})\} \quad (2)$$

Where $\boldsymbol{c}_j \in \Omega_i, v_j \in \mathcal{L}, \mathcal{L} = \{S_2, S_1, L_5, L_4, L_3, L_2, L_1, T_{12}, T_{11}, T_{10}\}$. We reduce the set of the spinal levels to T10, since the dataset that we used for the experiments does not capture any vertebrae at higher spinal levels. Finally, it is assumed that the field of view covers at least the S1 – L1 region.

2.1 Localization of Vertebral Bodies and Estimation of their Pose

The localization of the lumbar VBs is performed in 4 steps, summarized in **Fig. 1**.

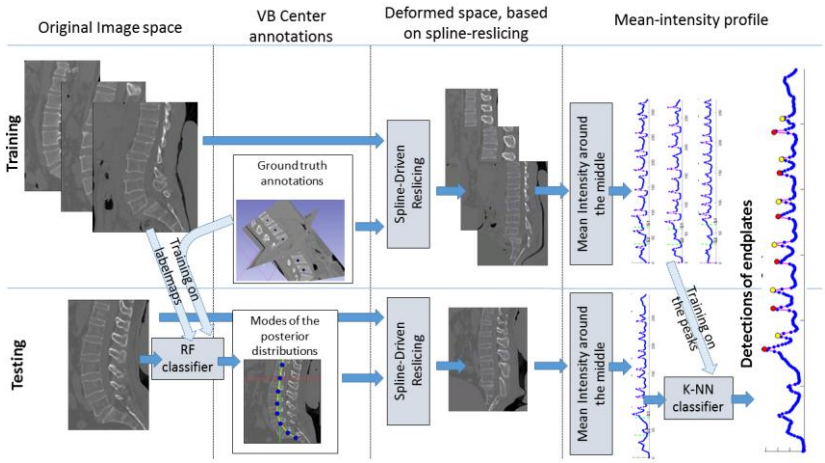

Fig. 1. A flowchart of the steps of proposed pipeline for the localization of the lumbar VBs. From left to right: a) A first-level detection of the VB centers is performed using the method of [7]; b) The original image is resliced along the curve that passes through the first-level detections; c) A mean-intensity profile is calculated along the axial center of the resliced image and the peaks of the mean-intensity profile that correspond to endplate locations are identified using a k-Nearest Neighbors classifier.

Firstly, a first-level detection of the centers is performed using the method proposed in [7]. This method employs a random forest classifier and in the present work it is used for a first-level detection of the VB centers of the levels $\mathcal{L}$. At training time, it constructs a label-map for every training image, using the ground-truth annotations of the VB centers. A random forest multi-label classifier is trained on the label-maps, using as features the mean intensity of displayed boxes (Haar-like features). At testing time, the generated probability map for every vertebral level is assumed to follow a normal distribution. The mode of the distribution for every generated probability map separately is retrieved with the Mean-Shift mode-seeking algorithm ([8]).

Secondly, the original CT image is resliced along the curve that passes through a set of VB center locations by performing an image deformation known as *Curved Planar*

Reformation ([9]). At training time, these locations are the ground truth VB center annotations whereas at testing time they are the first-level VB center detections. The reslicing is carried out using the method of [10], which firstly calculates a B-spline that passes through the first-level detections and then constructs a Local Coordinate System (LCS) on every point of the B-spline. In the rest of this paper, the resulting deformed image will be referred to as the *spline-unwrapped image*.

Thirdly, for every slice of the spline-unwrapped image, the mean intensity of a region around the middle point of the slice is calculated. The result of this operation is a univariate signal, referred to as the *mean-intensity profile* (shown in the top plot of **Fig. 2**). For the training phase, it is denoted with $s_i: O_i \subset \mathbb{N} \to \mathbb{R}$. For a testing image, it is denoted with s_T.

Lastly, the locations of VB centers are inferred from the positions of the vertebral endplates in the mean-intensity profile, using an approach very similar to those of [11] and [12]. Unlike [11], we do not attempt to detect periodic patterns in the mean intensity profile but we just locate its local maxima. As in [12], the basic observation for the detection of the endplates is that their locations correspond to local maxima (peaks) in the mean-intensity profile. Unlike [12], we do not make any assumptions concerning the orthogonal symmetry of the vertebrae in order to fine tune the VB center estimations and we just average the locations of the bottom and top endplates. Furthermore, we attempt to add robustness to the identification of the peaks that correspond to endplates by training a k-Nearest Neighbors classifier specifically for this task. In detail:

At training time, for every mean-intensity profile s_i, we locate the positions of those peaks $P_i = \{p_1, \cdots, p_{q_i}\}, p_1 \in O_i$ which are anatomically superior to the annotation $\boldsymbol{c}_{S_1}$. Hence, S1 is used as an anchor vertebra. For every peak position p_ξ we compute three simple features: (a) the value of the peak $s_i(p_\xi)$; (b) its left prominence e_ξ and (c) its right prominence E_ξ. The left prominence is defined as:

$$e_\xi = \max\{s_i(p_\xi) - s_i(\rho), \rho \in O_i,\ s_i \nearrow [\rho, p_\xi]\} \tag{3}$$

Where the $\nearrow$ denotes that s_i is increasing in the specified interval. The right prominence E_ξ is defined symmetrically. A binary label is provided for every peak, marking whether it corresponds to an endplate position or not. A k-Nearest Neighbors classifier is fit to this training set.

At testing time, the peaks of mean-intensity profile s_T after the S1 first-level detection are identified (**Fig. 1**a) and the 3 features are computed as in training. The trained k-Nearest Neighbors classifier classifies these peaks as corresponding to endplates or not (**Fig. 1**b). Finally, the estimates for the centers of the lumbar VBs are given by simply averaging the endplate positions. The pose of every vertebra is given by the LCS (computed at the second step) of the point of the B-spline which is closest to the VB center estimation.

2.2 Localization of Vertebral Landmarks

The objective of the second module is to locate a given number of vertebral landmarks on each level of the lumbar spine separately. We are interested in the lumbar

spinal levels $\mathcal{L}' = \{L_5, L_4, L_3, L_2, L_1\}$. Let $v \in \mathcal{L}'$ be one such level. For simplicity, it is assumed that same number $M_v = M$ of landmarks is desired to be found on all the levels. Therefore, it is assumed that the annotations B_i^v of the M landmarks of the vertebra at level v of every training image I_i are available. This ordered set of annotations is denoted as:

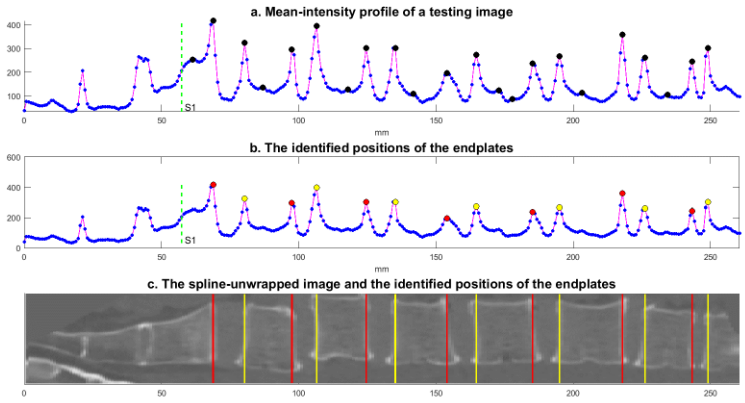

Fig. 2. The mean-intensity profile and the detection of the endplates. Top: The mean-intensity profile of a testing image with all the local maxima (peaks) in black dots; Middle: The output of the k-Nearest Neighbors classifier, which classifies the peaks in "endplate" and "non-endplate". Bottom: The same endplate positions, on the spline-unwrapped image. The prediction of the VB center is the average position of the bottom and top endplate on every lumbar level.

$$B_i^v = (\mathbf{c}_1^v, \cdots, \mathbf{c}_M^v), \mathbf{c}_j \in \Omega_i \tag{4}$$

The VB center annotations for the lumbar region of Eq. 2 are incorporated in the ordered sets B_i as its first elements, i.e. $\mathbf{c}_1^v = \mathbf{c}_v$ for all the levels, where $\mathbf{c}_v$ is annotation for the VB center.

Hence, given a testing image I_T, the task is to localize the M landmarks on each level $v \in \mathcal{L}'$. This is accomplished with two layers of random forest regressors, combined in an Auto-context fashion ([5]). The next two sections describe these two layers.

First Layer: Multi-Landmark Localization Using Appearance Features

The first layer employs a traditional multi-landmark regression-based method. One random forest regressor is trained for each lumbar level. Since each of these regressors is being utilized independently of all the others, let's assume that interest is on a specific level $v \in \mathcal{L}'$.

At training time, the images I_i are rotated according to the poses of the vertebrae at level v, so that all the vertebrae at level v are aligned, and a ROI is constructed around the VB center. The training set is sampled from all the ROIs. For every training sample,

the displacements to the M landmarks are computed, hence it is paired with $3 * M$ continuous values. A random forest is trained to regress these displacements. The traditional Haar-like features are used (as in [6,7,13]), which are based on the mean intensity value of randomly displayed boxes. The following feature types are considered:

$$f(\boldsymbol{x}; B_1, B_2, \boldsymbol{o}_1, \boldsymbol{o}_2, s) = \frac{\sum_{\boldsymbol{y} \in B_1} I(\boldsymbol{x}+\boldsymbol{o}_1+\boldsymbol{y})}{|B_1|} - s \frac{\sum_{\boldsymbol{y} \in B_2} I(\boldsymbol{x}+\boldsymbol{o}_2+\boldsymbol{y})}{|B_2|} \quad (5)$$

Where $s \in \{0,1\}$, $\boldsymbol{B}_1, \boldsymbol{B}_2$ are the sizes of two 3D boxes and $\boldsymbol{o}_1, \boldsymbol{o}_2$ are 3D offsets. A specific number of these features is sampled at the beginning of the random forest training and the parameters of the features are sampled uniformly from an interval of allowed values. In each leaf node of the decision trees of the trained random forest, two vectors of dimension $3 * M$ are stored: The mean displacements of the training samples that arrived on this leaf and their variance along every dimension.

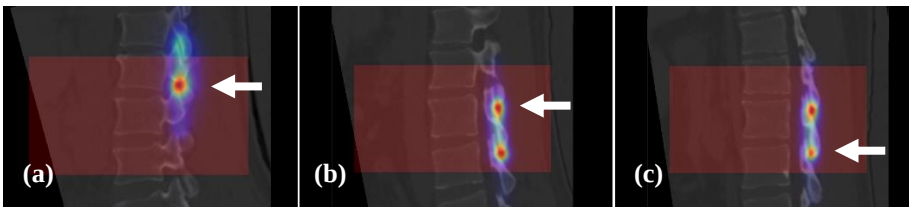

Fig. 3. Illustration of the problem with the single regressor approach on 3 testing images. The semi-transparent red layer represents the ROI testing region and the blue-red colormap the vote maps. All the testing images are rotated around the respective vertebra of interest, L1 for (a) and (c) and L2 for (b). The white arrow point to the ground-truth location of the corresponding landmark. On (a), the vote map retains its maximum value around the correct location. On (b), while the vote map still has a higher value around the correct location, it can no longer be considered unimodal. On (c), the problem is even clearer, as the mode with the highest value does not correspond to the ground truth location.

At testing time, the first module estimates the VB center of I_T at level v and the relevant pose. Then, I_T is aligned according the detected pose, a ROI is generated around the detected VB center and a testing set is sampled from inside this ROI. In a traditional single-layer approach, every testing sample would be parsed by every tree of the forest and it would cast M votes for the locations of each of the M landmarks. The aggregation of the votes from all the testing samples results in M maps, which in this work will be referred to as *vote maps*. The location of the each of the M landmarks would be inferred from its vote map, via a mode-seeking algorithm.

A drawback of this approach is that the vote maps are not guaranteed to be unimodal. In fact, it is to be expected that the vote map will have a high value on not only the target landmark at level v but possibly on the homologous location of neighboring vertebra with similar appearance. This is partially addressed by the fact that only a ROI around the detected VB center of the testing image is considered. However, the problem is not fully eliminated, since it is not possible to know in advance how large this ROI should be. This is illustrated in **Fig. 3**, where sagittal slices from 3 different testing

cases are depicted. The problem is most apparent in **Fig. 3**(c), where the maximum of the vote map occurs on the incorrect mode. A mode-seeking algorithm, without any additional post-processing, would fail in that case.

Second Layer: Addition of Context Features

The problem of the concurrent appearance of modes on neighboring spinal levels is addressed by the addition of context features. These context features are similar to the ones introduced in [6], where context information is extracted from the distance maps. In [6] one random forest regressor is constructed for every landmark. This would not scale well to the current task, as $|\mathcal{L}'| * M$ regressors would have to be trained on every layer. Therefore, the context features are used here by a multi-landmark regressor.

At training time, every tree of the first layer makes 3 predictions (one for each spatial dimension) for the displacement of each training sample to each of the M landmarks. The mean value of Euclidean distance of these 3 predictions over all the trees is called a *distance map*. At this point, an important decision to be made is which part of the training image should be considered for the computation of the distance maps. A simple choice is to use the same ROI as the one used for the training of the first layer. However, such a setup will bias the second layer into expecting that the VB center lies exactly at the center of the sampling ROI. In order to remove this bias, a modification to the standard Auto-context framework is introduced. For every ground truth annotation $\mathbf{c}_1^v$ of the VB centers, W randomly displaced locations are generated:

$$\tilde{\mathbf{c}}_{1,w}^v = \mathbf{c}_1^v + \boldsymbol{d}_w, \boldsymbol{d}_w \in [-d, d]^3, w \in \{1, \cdots, W\} \tag{6}$$

The displacements $\boldsymbol{d}_w$ are sampled randomly from the space $[-d, d]^3$. Then, W ROIs $\mathcal{R}_w \subseteq \Omega_w, w \in \{1, \cdots, W\}$ around each $\tilde{\mathbf{c}}_{1,w}^v$ are generated. The regions $\mathcal{R}_w$ of the training image I_i are parsed by the random forest of the first layer in order to compute the $W * M$ distance maps $\mathrm{D}_{i,w}^m: \Omega_i \to \mathbb{R}^+$. The value of distance maps $\mathrm{D}_{i,w}^m$ outside of $\mathcal{R}_w$ is set to a fixed, large value.

For the training of the second layer, each training image I_i is taken into account W times, each time paired with the M distance maps $\mathrm{D}_{i,w}^m$. As in the first layer, a pool of Haar-like features is sampled before the training of the forest starts, with the difference that the intensity image $I(\cdot)$ of Eq. 5 can be replaced by one of the $\mathrm{D}_{i,w}^m$. When this happens, the resulting Haar-like feature is able to capture context information from the distance maps of the first layer (context feature). The total number of context features is set beforehand as a hyperparameter.

At testing time, the distance maps of the testing image are generated by the first layer, using the same testing ROI as in the first layer. The testing image is paired with the generated distance maps and it is passed to the second layer, so that both appearance features (computed on the original testing image) and context features (computed on the distance maps) can be calculated. The testing pipeline proceeds with the computation of the vote maps. For every testing sample, the vote generated by each tree of the forest for each landmark is taken into account separately, provided that the variance of the displacement is below a certain threshold. As it is illustrated in the second row of **Fig. 4**, the resulting vote map is unimodal. The mode of every vote map is estimated

using the mean-shift algorithm ([8]). Finally, the estimated modes are rotated back to the original image space in order to provide the localization of the landmarks.

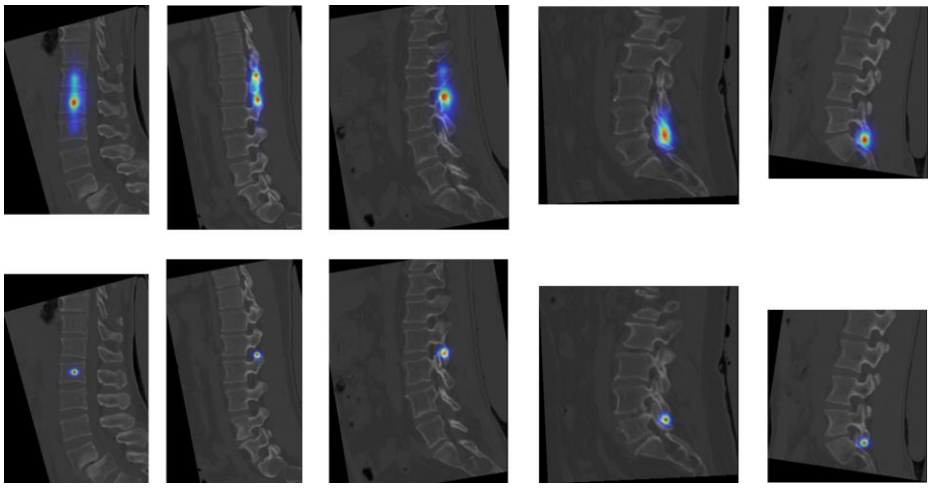

Fig. 4. Qualitative comparison of the voting maps of a single random forest layer (top row) with the voting maps after the second layer of the proposed method (bottom row). From left to right: Sagittal cuts of different testing images for levels L1 to L5, respectively. Notice that the images have been automatically aligned around the vertebra of interest. In can be observed that the voting map is more concentrated in the two-layered approach and it is has exactly one mode around the correct landmark location.

3 Experiments and Results

3.1 Dataset and Experimental Setup

The proposed methods are evaluated on a dataset of 28 CT images. The intra-slice slice spacing is in the 0.29 – 0.42 mm range, the inter-slice spacing is 0.7 mm and the slice size is 512x512. All of the images capture at least the S1 – L1 levels, which is typical for scans of the lumbar spine. The thoracic region is captured up to the T10 level some cases. No implants are presented in any of the images. There are cases with mild scoliosis, osteophytes and fractures vertebrae. For every lumbar vertebra, 5 manual annotations are made: The center of the VB and the tips of the four inferior articular processes. There are four inferior articular process on a typical lumbar vertebra: a bottom-left, a bottom-right, a top-right and a top-left. We will refer to their tips as landmarks A, B, C and D respectively. Sagittal and coronal view of two example annotations for landmarks A and B are shown on **Fig. 5**(a) and **Fig. 5**(b).

All of the images are resampled to an isotropic spacing of 1x1x1 mm. We randomly select 20 images to be used for the training phase of the proposed methods. The held-out 8 images will be used for evaluation.

The hyperparameters of the proposed methods are set through a leave-one-out cross-validation iteration on the training set. The K parameter of the k-Nearest Neighbors classifier of the first module (localization of VBs) is set to 15. For the second module (localization of vertebral landmarks) the hyperparameters are as following: For the random forest of the first layer, 50 trees are trained, the size of the feature pool is 10000 and on every node the search space is 200 features. The size of the ROI around every vertebra, during both training and testing, is 120x150x80 mm. The training set of every tree is a random 1% subset of all the voxels inside the ROIs of the images. At testing time, all of voxels inside the ROIs are used. For the second layer, the parameters W, d of Eq. 6 are set to 5 and 8 mm respectively. 50 trees are trained with a depth of 25. The size of the feature pool is 11000 features: 10000 features plus exactly 200 from the each of the 5 distance maps. The sampling of the training samples is again 1%. At testing time, all the voxels inside the ROI are tested, but only the votes with a predicted variance of less than 15 mm in every spatial dimension (from each tree independently) are taken into account for the construction of the vote maps.

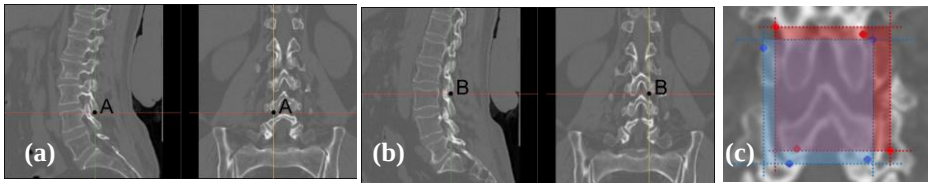

Fig. 5. Left: the sagittal and the coronal views of the annotation of the tip of the bottom-left inferior articular process (landmark A) of an L3 vertebra. Middle: The sagittal and coronal views of the annotation of a bottom-right inferior articular process (landmark B) of an L2 vertebra. Right: The 3D bounding boxes of the manual annotations for the inferior articular process (blue dots) and their detections (red dots). The localization errors have been exaggerated.

3.2 Evaluation

For the evaluation of the first module, two metrics are used: (a) the **rate of successful detections** and (b) the displacements to the manual annotations of the VB centers of the lumbar spine (**localization error**). A VB center detection is considered successful when it lies within 10 mm from the respective manual annotation. The detailed evaluation for every lumbar spinal level is presented on **Table 1**, where the rate of successful detections is labeled as "Id. Rate". All of the lumbar VB of the 8 testing images are detected successfully. The mean localization error is 3.2 mm, with a standard deviation of 2.0 mm and a median value of 2.8 mm. The evaluation of the first-level detections obtained with the method of [7] is also presented on **Table 1**.

For the evaluation of the second module, the localization error metric is also used. The localization errors for each lumbar level are presented in **Table 2 - Table 6**, along with the rate of the detections with a localization error of less than 6 mm. Overall, the proposed method achieves a mean localization error of 3.0 mm, with a 1.6 mm standard deviation and a median value of 2.7 mm. 95.4% of the detections have a localization error of below 6 mm. Regarding the training of the second layer, we experimented with

removing the randomly displaced ROIs of Eq. 6 and train instead using ROIs centered around the VB centers. With that setup, the localization error increases to 3.4 ± 1.8 mm.

As an additional metric for the quality of the detections, their bounding boxes are also considered. In particular, the extreme locations of the 5 landmarks in the each of the 3 spatial dimensions define 6 bounding planes and therefore a 3D bounding box. Coronal projections of such bounding boxes are depicted in **Fig. 5**(c) for both the manual annotations (blue box) and the automatic detections (red box). The evaluation metric is the **Dice overlap coefficient** of the bounding box of the manual annotations and the bounding box of the detections. The achieved scores on this metric are presented in **Table 7**. The mean dice coefficient, across all the spinal levels, is 88.8%.

Table 1. Localization performance of the first module for the VB centers. A detection is considered successful if it lies within 10 mm from the manual annotation (Id. Rate). The mean, standard deviation and median of the localization errors are computed on the successful detections only. The first-level detections are the output of the method [7]. The endplate-based detections are the output of the first module. Expect for the rates, all the quantities are expressed in mm.

	First-level detections					Endplate-based detections				
	L1	L2	L3	L4	L5	L1	L2	L3	L4	L5
Id. Rate (%)	75.0	62.5	62.5	75.0	100	100	100	100	100	100
Mean	3.7	6.7	8.1	7.6	5.0	2.8	3.4	3.8	3.1	2.9
Std.	2.1	3.5	3.3	3.2	1.8	1.3	2.3	1.5	2.6	1.5
Median	3.4	6.1	6.9	7.6	5.2	2.4	2.4	4.0	1.7	2.7

Table 2. Localization errors of the second module for the 5 landmarks of the L1 level

L1-level Landmarks	VB Center	A	B	C	D	Overall
Loc. error < 6 mm (%)	100	100	100	100	100	100
Mean (mm)	1.8	2.8	3.0	2.6	3.4	2.7
Std. (mm)	0.6	1.4	1.3	1.1	1.5	1.3
Median (mm)	1.7	2.6	2.7	2.6	2.9	2.7

Table 3. Localization errors of the second module for the 5 landmarks of the L2 level

L2-level landmarks	VB Center	A	B	C	D	Overall
Loc. error < 6 mm (%)	100	100	100	100	100	100
Mean (mm)	2.4	2.2	2.1	3.1	2.6	2.5
Std. (mm)	1.1	0.8	0.8	1.2	1.4	1.1
Median (mm)	2.8	2.2	2.0	2.8	2.4	2.3

Table 4. Localization errors of the second module for the 5 landmarks of the L3 level

L3-level landmarks	VB Center	A	B	C	D	Overall
Loc. error < 6 mm (%)	100	100	87.5	87.5	100	95.0
Mean (mm)	2.7	3.4	3.7	3.6	2.8	3.2
Std. (mm)	0.9	1.6	1.5	1.9	0.9	1.5
Median (mm)	2.4	3.1	3.1	3.1	2.7	2.9

Table 5. Localization errors of the second module for the 5 landmarks of the L4 level. One testing case has been omitted because it was not possible to annotate all its articular processes due to a vertebral fracture. Hence, there are 7 testing images on this spinal level.

L4-level landmarks	VB Center	A	B	C	D	Overall
Loc. error < 6 mm (%)	100	87.5	87.5	87.5	87.5	88.6
Mean (mm)	1.9	4.0	3.4	3.5	3.4	3.2
Std. (mm)	1.0	1.7	2.3	1.9	2.9	2.2
Median (mm)	1.7	3.2	2.1	3.5	2.4	2.4

Table 6. Localization errors of the second module for the 5 landmarks of the L5 level

L5-level landmarks	VB Center	A	B	C	D	Overall
Loc. error < 6 mm (%)	100	75.0	100	100	100	92.5
Mean (mm)	2.7	4.5	3.4	3.4	2.9	3.4
Std. (mm)	1.6	1.6	1.3	1.3	1.1	1.5
Median (mm)	2.4	4.7	3.8	3.3	2.9	3.0

Table 7. Dice Coefficients for the bounding boxes from the 5 landmarks over each spinal level

Dice Coefficients	L1	L2	L3	L4	L5	Overall
Mean	0.87	0.90	0.90	0.90	0.87	0.89
Min.	0.80	0.86	0.87	0.80	0.82	0.80
Max.	0.94	0.95	0.96	0.95	0.92	0.96

4 Conclusion

The repetitive nature of the spine poses an additional difficult to the task of landmark localization, as neighboring vertebrae often have very similar appearance. However, a fully automatic method for localizing vertebral landmarks is highly desirable, as it can provide as a robust initialization step for model-based segmentation methods and it can facilitate the assessment of certain vertebral pathologies. In this work, a pipeline for the detection of lumbar vertebral landmarks is proposed. The proposed pipeline starts with the detection of VB centers and proceeds with the localization of landmarks on each lumbar level. For evaluation, the pipeline was applied for the localization of the VB centers and the inferior articular processes on a dataset of lumbar-focused CT images. The experimental results suggest that the proposed method can detect reliably the vertebral landmarks on all the levels of the lumbar spine. Even though in our experiments we focused on the articular processes, we expect that the proposed method can be applied for different vertebral landmarks as well, such as key endplate landmarks for the measurement of spondylolisthesis. In the future, we plan to explore such a direction. Future research also includes the more extensive evaluation of the proposed methods on larger datasets and the investigation of ways to improve the localization accuracy, for example by fine-tuning the detections in a multi-scale fashion and by introducing context features from different vertebral levels.

References

1. J. Reginster, "The prevalence and burden of arthritis," Rheumatology, vol. 41, no. suppl 1, pp. 3–6, 2002.
2. K. Knapp and G. Slabaugh, "Improving an Active Shape Model with Random Classification Forest for Segmentation of Cervical Vertebrae," presented at the Computational Methods and Clinical Applications for Spine Imaging: 4th International Workshop and Challenge, CSI 2016, Held in Conjunction with MICCAI 2016, Athens, Greece, October 17, 2016, Revised Selected Papers, 2017, vol. 10182, p. 3.
3. M. Roberts, T. Cootes, and J. Adams, "Automatic location of vertebrae on DXA images using random forest regression," Medical Image Computing and Computer-Assisted Intervention–MICCAI 2012, pp. 361–368, 2012.
4. S. Liao *et al.*, "Automatic lumbar spondylolisthesis measurement in ct images," *IEEE transactions on medical imaging*, vol. 35, no. 7, pp. 1658–1669, 2016.
5. Z. Tu, "Auto-context and its application to high-level vision tasks," presented at the Computer Vision and Pattern Recognition, 2008. CVPR 2008. IEEE Conference on, 2008, pp. 1–8.
6. Y. Gao and D. Shen, "Context-aware anatomical landmark detection: application to deformable model initialization in prostate CT images," presented at the International Workshop on Machine Learning in Medical Imaging, 2014, pp. 165–173
7. B. Glocker, D. Zikic, E. Konukoglu, D. R. Haynor, and A. Criminisi, "Vertebrae localization in pathological spine CT via dense classification from sparse annotations," presented at the International Conference on Medical Image Computing and Computer-Assisted Intervention, 2013, pp. 262–270.
8. Y. Cheng, "Mean shift, mode seeking, and clustering," IEEE transactions on pattern analysis and machine intelligence, vol. 17, no. 8, pp. 790–799, 1995.
9. A. Kanitsar, D. Fleischmann, R. Wegenkittl, P. Felkel, and M. E. Gröller, "CPR: curved planar reformation," presented at the Proceedings of the conference on Visualization'02, 2002, pp. 37–44.
10. Velut J., A Spline-Driven Image Slicer. Published in The VTK Journal. 2011.
11. D. Štern, B. Likar, F. Pernuš, and T. Vrtovec, "Automated detection of spinal centrelines, vertebral bodies and intervertebral discs in CT and MR images of lumbar spine," *Physics in medicine and biology*, vol. 55, no. 1, p. 247, 2009.
12. D. Forsberg, C. Lundström, M. Andersson, L. Vavruch, H. Tropp, and H. Knutsson, "Fully automatic measurements of axial vertebral rotation for assessment of spinal deformity in idiopathic scoliosis," *Physics in medicine and biology*, vol. 58, no. 6, p. 1775, 2013.
13. A. Criminisi et al., "Regression forests for efficient anatomy detection and localization in computed tomography scans," Medical image analysis, vol. 17, no. 8, pp. 1293–1303, 2013.