Spine-Rib Segmentation and Labeling via Hierarchical Matching and Rib-Guided Registration

Caiwen Jiang[1], Zhiming Cui[1], Dongming Wei[1], Yuhang Sun[1], Jiameng Liu[1], Jie Wei[1], Qun Chen[1,2], Dijia Wu[1,2], and Dinggang Shen[1,2(✉)]

[1] School of Biomedical Engineering, ShanghaiTech University, Shanghai, China
dgshen@shanghaitech.edu.cn

[2] Shanghai United Imaging Intelligence Co., Ltd., Shanghai, China

Abstract. Accurate segmentation and labeling of spine-rib are of great importance for clinical spine and rib diagnosis and treatment. In clinical applications, the spine-rib segmentation and labeling are often challenging, as the shape and appearance of vertebrae are complicated. Previous segmentation and labeling methods usually face considerable difficulties when coping with spine CT images with abnormal curvature spines and implanted metal. In this paper, we propose a multi-stage spine-rib segmentation and labeling method that can be applied to various spine-rib CT images. Our proposed method consists of three steps. First, a 3D U-Net is used to obtain a initial segmentation mask of the spine and rib. Then, the subject information, including gender, age, and the shape of the spine and rib, is used for hierarchically selecting the templates with similar physiological structures from the pre-constructed template library. Finally, the segmentation mask and label from the templates are transferred to the subject via rib-guided registration to achieve correction of the initial results. We evaluated the proposed method on a clinical dataset, and obtained significantly better and robust performance than the state-of-the-art method.

Keywords: Spine-rib segmentation and labeling · Hierarchical matching · Rib-guided registration · Templates

1 Introduction

Spine-rib segmentation and labeling are important for image-guided diagnosis, pre-operative planning, and post-operative evaluation [1,2]. In conventional clinical diagnosis, the doctor needs to determine the type of vertebrae and ribs based on experience, and then segment them slice by slice. Thus, manual segmentation and labeling of vertebrae and ribs in CT images is laborious and subjective [3]. Therefore, it is necessary to propose an automatic spine-rib segmentation and labeling algorithm to improve efficiency and reliability.

C. Lian et al. (Eds.): MLMI 2021, LNCS 12966, pp. 537–545, 2021.
https://doi.org/10.1007/978-3-030-87589-3_55

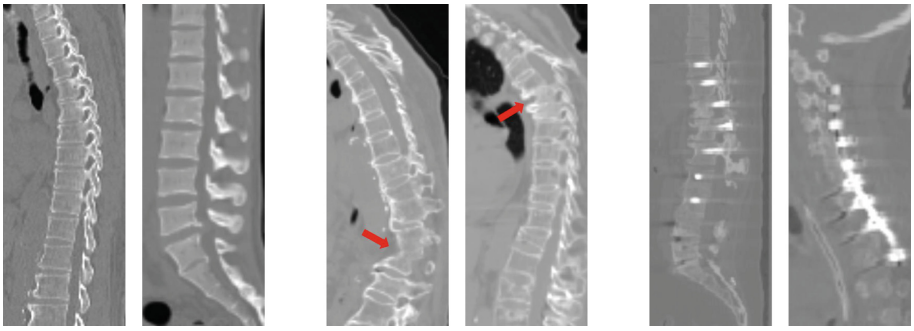


Fig. 1. Sagittal plane of six cases. (a) The FoV of images varies largely, and the appearance of adjacent vertebrae is very similar. (b) Abnormal curvature of the spine. (c) The surgical metal implants cause peculiar image artifacts.

Automatic spine-rib segmentation and labeling have been applied in various clinical applications, such as detection of vertebra and rib fractures [4], assessment of spinal deformities [5], and computer-assisted surgical interventions [6]. But there are still many challenges in the clinical stage to design an automated spine-rib segmentation and labeling algorithm. As shown in Fig. 1(a), the field of view (FoV) of spine CT images varies largely, and the appearance of adjacent vertebrae is too similar to distinguish. Various pathological circumstances, including scoliosis, vertebra fractures, and lumbarization, increase the difficulty of vertebrae identification, as shown in Fig. 1(b). Moreover, as shown in Fig. 1(c), the presence of surgical metal implants usually causes severe blurring of the vertebral boundary.

Recently, many approaches have been proposed to solve the problems mentioned above, which can be divided into three categories. The first combines machine learning and statistical models [7–9], which are robust to be applied to various spine-rib CT images. However, this category of the method is hard to be used in the application stage, due to the requirement of hand-crafted image features. The second category is based on multi-stage neural networks [10–14], which can effectively extract global context information of vertebrae and ribs to solve the arbitrary FoV problem. But this category of methods can not robustly handle pathological or abnormal spinal images. The third category of methods is based on template matching [15–17]. However, the diversity of templates used in previous methods are limited and cannot cover the complex situation.

In this paper, to address the above-mentioned limitations, we present an accurate and stable spine-rib segmentation and labeling method via hierarchical matching and rib-guided registration. In the first stage, we construct a representative template library by collecting numerous spine-rib CT images. Then given a testing CT image, we first use a 3D U-Net to obtain the semantic segmentation of vertebrae and ribs. In addition, we use information such as gender, age, spine, and rib shape to perform hierarchical matching, and select templates that have similar structures to the input object. Finally, we obtain the corrected spine and

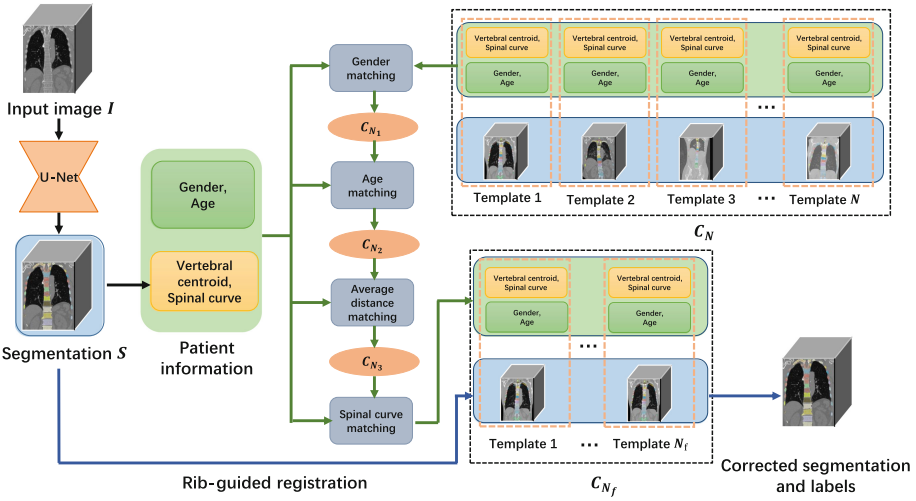


Fig. 2. The overall architecture of the proposed method, where C_N, C_{N_1}, C_{N_2}, C_{N_3} and C_{N_f} denote the template set, and N, N_1, N_2, N_3 and N_f denote the numbers of templates where $(N_f \leqslant N_3 \leqslant N_2 \leqslant N_1 \leqslant N)$.

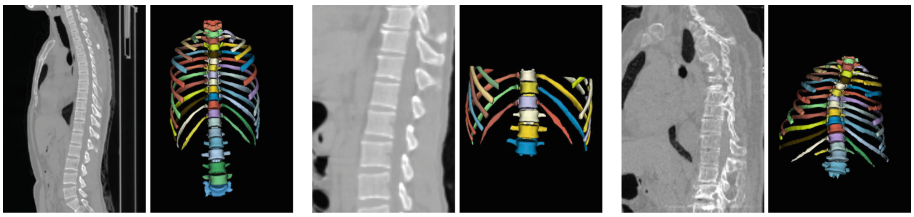
rib segmentation masks and labels by transferring the annotation from selected templates to the subject via rib-guided registration. We extensively evaluate our method on a clinical dataset with various pathological or abnormal spinal CT images, and results are significantly better than the state-of-the-art methods.

2 Method

As shown in Fig. 2, our proposed method consists of three stages. First, we use a 3D U-Net to obtain initial semantic segmentation S of the vertebrae and ribs, from which spinal structure information such as centroids of labels (vertebrae) and spinal curve can be directly obtained. Then, the spine structure information and physiological information such as gender and age are used to select a template set C_{N_f}, have in which the vertebra and rib from template library C_{N_f} has the most similar structures with S by hierarchical matching. Finally, through rib-guided registration, we register vertebra and rib segmentation masks in C_{N_f} to S for obtaining the corrected segmentation results and labels.

2.1 Template Library Construction

We construct the template library C_N based on a representative clinical dataset, where three examples are shown in Fig. 3. The templates in C_N are selected by three steps. First, the templates are selected to cover all genders, ages, and physiques. Then, both normal and abnormal subjects (with abnormal curvature spine and metal implants) are included in the library to maintain the diversity.



(a) (b) (c)

Fig. 3. Three typical examples of the template library C_N, including (a) image of the whole view, (b) image of limited view, and (c) image of the pathological spine.

Finally, the segmentation and label of vertebrae and ribs are manually annotated by doctors.

2.2 Hierarchical Matching

In this section, we first use gender and age to perform initial selection to obtain C_{N_1}. Then, the average distance between the vertebrae is used to select templates that have the similar physique to the input subject. Finally, we select C_f from C_N according to the similarity of spinal curves.

Gender and Age Screening. First, we use gender screening to get C_{N_1} from C_N, where the selected templates have the same gender as input image I. Second, the templates C_{N_2} within three years gap are screened from C_{N_1}. Note that it is unnecessary to filter age precisely, as the spine structure should be consistent within several years.

Average Distance Screening. As individuals of the same gender and age still have different physique, we can select C_{N_3} with the similar physique as I from C_{N_2} by average distance matching. First, we approximately calculate the centroids of labels by averaging the coordinates corresponding to same vertebrae. Then, the length of the entire spinal segment can be calculated by using the coordinates of the starting centroid and the ending centroid in the spine. Finally, the length of the entire spinal segment is divided by the number of vertebrae to obtain the average distance d_p.

The spine sequence in S and templates C_{N_2} may be different, therefore, we need to select the templates containing the same corresponding spine segment as S, and then calculate the template average distance d_t at the same spine segment. Through extensive experiments, we find that, although the labels of vertebrae in S are unreliable, the spine segment in S can be approximately determined by these labels. By comparing d_p and d_t, we can perform further selection to obtain C_{N_3}.

Spinal Curve Screening. Further, we select C_f with similar spinal curvatures as I from C_3 by spinal curve matching. First, the spinal curve can be obtained by performing cubic spline interpolations based on the extracted vertebral centroid positions. The Chamfer distance [13] is calculated from the rigidly aligned spinal curves of I to spinal curves of templates in C_3, and N_f templates with the shortest distance are selected to form C_{N_f}.

2.3 Rib-Guided Registration

After hierarchical matching, the spine physiological structures of templates in C_{N_f} are similar to I. Then, we perform registration between S and templates in C_{N_f} for obtaining corrected segmentation masks and labels. However, considering the shape of adjacent vertebrae are similar and difficult to distinguish, direct registration between S and templates in C_f may lead to misalignment. To tackle this issue, we find that the ribs connected to the vertebrae are easily recognized due to the clear difference in rib length. Therefore, the rib is used to guide the registration with two stages. First, the ribs segmentation masks in S and C_{N_f} are aligned. Then, the registration of vertebrae and ribs is jointly performed between S and C_{N_f}.

Note that the semantic segmentation of I is obtained through a pretrained 3D U-Net. Therefore, we can obtain the segmentation and labeling (S_{rib}) of each rib from S according to the rib length. For the template, we can perform the same operation to obtain segmentation masks and labels of ribs T_{rib}. The segmentation masks of vertebrae and ribs in the templates are denoted as T. We register T_{rib} to S_{rib} for obtaining the deformation field ϕ_{rib}. The registration loss is defined over the rib masks as follows:

$$\mathcal{L}_{rib} = \mathcal{L}_{sim}(T_{rib}(\phi_{rib}), S_{rib}) + \lambda \mathcal{L}_{smooth}(\phi_{rib}). \tag{1}$$

where $\mathcal{L}_{sim}(\cdot, \cdot)$ measures the mean square error (MSE) between $S_{rib}(\phi_{rib})$ and T_{rib}. $\mathcal{L}_{smooth}(\cdot)$ is a regularization term to constrain the deformation field to be smooth, and λ is the balance weight for the regularization.

Through registration between S_{rib} and T_{rib}, the corresponding vertebrae in S and T have been roughly aligned. Then, performing registration between S and T can effectively avoid the interference caused by the similarity of adjacent vertebrae. The deformation field obtained is denoted as ϕ_{rib_vert}. Then the whole rib-guided registration process can be formulated by the following equation:

$$\mathcal{L}_{rib\_guided} = \mathcal{L}_{sim}(T(\phi_{rib} \circ \phi_{rib\_vert}), S) + \lambda \mathcal{L}_{smooth}(\phi_{rib} \circ \phi_{rib\_vert}). \tag{2}$$

There are N_f templates in C_{N_f}, and the rib-guided registration is performed N_f times. Using MSE as the evaluation criterion, the template with the minimum MSE is selected as the final reference template. The corrected label can obtain directly from the reference template, and the corrected segmentation masks by $T(\phi_{rib} \circ \phi_{rib\_vert})$.

3 Experiments

3.1 Dataset and Evaluation Metrics

We collect a clinical dataset that includes CT scans of 1500 patients, of which 802 are male and 698 are female, with the ages distribution ranging from 8 to 99. The segmentation masks and labels of the vertebrae and ribs are manually annotated by doctors. There are 800 scans used for training the segmentation network, 600 scans used for template library construction, and 100 scans used for testing.

To quantitatively evaluate the performance of our proposed method, the segmentation performance is evaluated using both Dice and Hausdorff distance (HD). The labeling performance is evaluated by I_{acc} and S_{acc}, where I_{acc} is the percentage of vertebrae that are assigned with the correct label, and S_{acc} is the percentage of whole scans with correct labels for all vertebrae labels.

3.2 Implementation Details

In training U-Net for segmentation, the Adam optimizer is used with an initial learning rate $\lambda = 0.01$ and batch size = 2. The intensity is first normalized to $[0, 1]$. And all the images are resampled to be $256 \times 256 \times 256$ and have $2 \times 2 \times 2\text{mm}^3$ voxel size. All experiments are conducted on two NVIDIA Tesla V100 GPUs using the PyTorch platform.

During the hierarchical matching phase, we use 3 years as the gap for age matching, and match the average distance based on condition $|d_p \pm 2mm| \leqslant d_t$. Then, the Chamfer distance is calculated after aligning different spinal curves by coherent point drift (CPD) rigid registration [18], and the 3 templates with the highest overlap rate are selected to form C_{N_f}. In the stage of rib-guided registration, registration is achieved by affine transformation in ANTs [19] library.

3.3 Ablation Studies

In this section, we conduct extensive experiments to validate the effectiveness of key steps in our proposed framework. We compared our method with the segmentation masks of vertebra and rib using the same 3D U-Net network. The quantitative results are shown in Table 1. It can be found that the performance of segmentation and labeling is significantly improved by our proposed method, i.e., 9.36% improvements of Dice score and also 15.25 improvement of I_{acc}. This proves the effectiveness of rib-guided registration.

3.4 Evaluation and Comparison

Labeling results of our proposed method on five challenging cases are shown in Fig. 4, including various FoVs, metal artifacts, and abnormal curvature. Our proposed method achieves accurate labeling results even on the images with various FoVs and metal artifacts, (see Figs. 4(a) and 4(b)).

Table 1. Quantitative results of ablation studies.

Methods	Segmentation		Labeling	
	Dice (%)	HD (mm)	I_{acc} (%)	S_{acc} (%)
U-Net	81.12 ± 2.23	13.27 ± 4.21	82.11	79.00
Ours	$\mathbf{90.51 \pm 1.54}$	$\mathbf{5.79 \pm 1.27}$	**97.36**	**95.00**

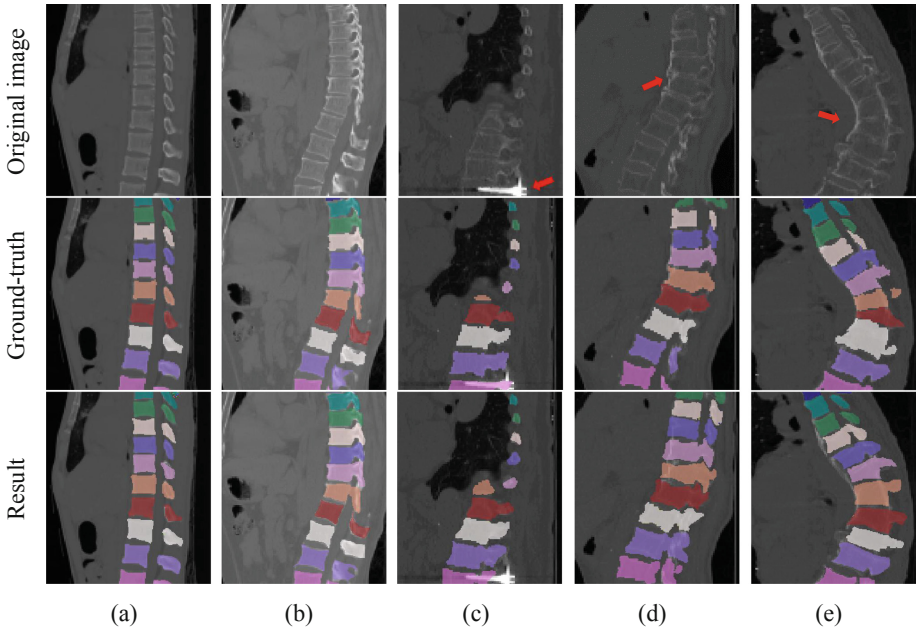


Fig. 4. Labeling results of our proposed method on five challenging cases, including (a)(b) limited FoV, (c) metal artifacts, and (d)(e) abnormal curvature. The top row shows are original images, the middle row shows ground-truth, and the bottom row shows our labeling results.

Although our method achieves reliable performance, there are still some limitations of our framework. For example, our proposed method obtains over-segmentation results over the boundaries of rib in Fig. 4(e). The main reason is that the template library does not contain a template very similar to this subject. To address this issue, we plan to collect more representative subjects to enrich the template library in the future.

For quantitative comparison, we compare our proposed method with three related methods proposed by Chen et al. [20], Lessmann et al. [21], and Payer et al. [22], respectively. The results of spine-rib segmentation and labeling from our proposed method and the three comparison methods on our dataset are shown in Table 2. From the results, our proposed method achieves the best performance in both segmentation and labeling tasks.

Table 2. Comparison on segmentation and labeling results

Methods	Segmentation		Labeling	
	Dice (%)	HD (mm)	I_{acc} (%)	S_{acc} (%)
Chen et al. [20]	82.56 ± 2.23	12.27 ± 4.21	85.23	83.00
Lessmann et al. [21]	83.42 ± 1.87	9.87 ± 2.56	88.45	86.00
Payer et al. [22]	88.34 ± 1.67	7.34 ± 1.72	92.83	90.00
Ours	$\mathbf{90.51 \pm 1.54}$	$\mathbf{5.79 \pm 1.27}$	**97.36**	**95.00**

4 Conclusion

In this paper, we have presented a novel method for spine-rib segmentation and labeling. Our method first utilizes a deep learning model to obtain the initial segmentation and label results. Then, a representative spine-rib template library is conducted to match the testing subject, for considering possible mistakes in spine CT images with complicated appearance. Through extensive experiments, our method shows significant performance improvements in spine-rib segmentation and labeling.

References

1. Ben Ayed, I., Punithakumar, K., Minhas, R., Joshi, R., Garvin, G.J.: Vertebral body segmentation in MRI via convex relaxation and distribution matching. In: Ayache, N., Delingette, H., Golland, P., Mori, K. (eds.) MICCAI 2012. LNCS, vol. 7510, pp. 520–527. Springer, Heidelberg (2012). https://doi.org/10.1007/978-3-642-33415-3_64
2. Lecron, F., Boisvert, J., Mahmoudi, S., Labelle, H., Benjelloun, M.: Fast 3D spine reconstruction of postoperative patients using a multilevel statistical model. In: Ayache, N., Delingette, H., Golland, P., Mori, K. (eds.) MICCAI 2012. LNCS, vol. 7511, pp. 446–453. Springer, Heidelberg (2012). https://doi.org/10.1007/978-3-642-33418-4_55
3. Burns, J.E.: Imaging of the spine: a medical and physical perspective. In: Li, S., Yao, J. (eds.) Spinal Imaging and Image Analysis. LNCVB, vol. 18, pp. 3–29. Springer, Cham (2015). https://doi.org/10.1007/978-3-319-12508-4_1
4. Yao, J., Burns, J.E., Munoz, H., Summers, R.M.: Detection of vertebral body fractures based on cortical shell unwrapping. In: Ayache, N., Delingette, H., Golland, P., Mori, K. (eds.) MICCAI 2012. LNCS, vol. 7512, pp. 509–516. Springer, Heidelberg (2012). https://doi.org/10.1007/978-3-642-33454-2_63
5. Forsberg, D., Lundström, C., Andersson, M., Vavruch, L., Tropp, H., Knutsson, H.: Fully automatic measurements of axial vertebral rotation for assessment of spinal deformity in idiopathic scoliosis. Phys. Med. Biol. **58**(6), 1775 (2013)
6. Knez, D., Likar, B., Pernuš, F., Vrtovec, T.: Computer-assisted screw size and insertion trajectory planning for pedicle screw placement surgery. IEEE Trans. Med. Imaging **35**(6), 1420–1430 (2016)
7. Glocker, B., Feulner, J., Criminisi, A., Haynor, D.R., Konukoglu, E.: Automatic localization and identification of vertebrae in arbitrary field-of-view CT scans. In:

Ayache, N., Delingette, H., Golland, P., Mori, K. (eds.) MICCAI 2012. LNCS, vol. 7512, pp. 590–598. Springer, Heidelberg (2012). https://doi.org/10.1007/978-3-642-33454-2_73
8. Glocker, B., Zikic, D., Konukoglu, E., Haynor, D.R., Criminisi, A.: Vertebrae localization in pathological spine CT via dense classification from sparse annotations. In: Mori, K., Sakuma, I., Sato, Y., Barillot, C., Navab, N. (eds.) MICCAI 2013. LNCS, vol. 8150, pp. 262–270. Springer, Heidelberg (2013). https://doi.org/10.1007/978-3-642-40763-5_33
9. Zhan, Y., Maneesh, D., Harder, M., Zhou, X.S.: Robust MR spine detection using hierarchical learning and local articulated model. In: Ayache, N., Delingette, H., Golland, P., Mori, K. (eds.) MICCAI 2012. LNCS, vol. 7510, pp. 141–148. Springer, Heidelberg (2012). https://doi.org/10.1007/978-3-642-33415-3_18
10. Chen, H., et al.: Automatic localization and identification of vertebrae in Spine CT via a joint learning model with deep neural networks. In: Navab, N., Hornegger, J., Wells, W.M., Frangi, A.F. (eds.) MICCAI 2015. LNCS, vol. 9349, pp. 515–522. Springer, Cham (2015). https://doi.org/10.1007/978-3-319-24553-9_63
11. Yang, D., et al.: Deep image-to-image recurrent network with shape basis for automatic vertebra labeling in large-scale 3D CT volumes, 30 July 2019. US Patent 10,366,491
12. Yang, D., et al.: Automatic vertebra labeling in large-scale 3D CT using deep image-to-image network with message passing and sparsity regularization. In: Niethammer, M., et al. (eds.) IPMI 2017. LNCS, vol. 10265, pp. 633–644. Springer, Cham (2017). https://doi.org/10.1007/978-3-319-59050-9_50
13. Cui, Z., et al.: TSegNet: an efficient and accurate tooth segmentation network on 3D dental model. Med. Image Anal. **69**, 101949 (2021)
14. Alomari, R.S., Ghosh, S., Koh, J., Chaudhary, V.: Vertebral column localization, labeling, and segmentation. In: Li, S., Yao, J. (eds.) Spinal Imaging and Image Analysis. LNCVB, vol. 18, pp. 193–229. Springer, Cham (2015). https://doi.org/10.1007/978-3-319-12508-4_7
15. Ullmann, E., Paquette, J.F.P., Thong, W.E., Cohen-Adad, J.: Automatic labeling of vertebral levels using a robust template-based approach. Int. J. Biomed. Imaging **2014**, 719520 (2014)
16. Larhmam, M.A., Benjelloun, M., Mahmoudi, S.: Vertebra identification using template matching modelmp and K-means clustering. Int. J. Comput. Assist. Radiol. Surg. **9**(2), 177–187 (2013)
17. Wu, D.: A learning based deformable template matching method for automatic rib centerline extraction and labeling in CT images. In: 2012 IEEE Conference on Computer Vision and Pattern Recognition, pp. 980–987. IEEE (2012)
18. Myronenko, A., Song, X.: Point set registration: coherent point drift. IEEE Trans. Pattern Anal. Mach. Intell. **32**(12), 2262–2275 (2010)
19. Avants, B.B., Tustison, N., Song, G.: Advanced normalization tools (ants). Insight J. **2**(365), 1–35 (2009)
20. Sekuboyina, A., et al.: Verse: a vertebrae labelling and segmentation benchmark. arXiv preprint arXiv:2001.09193 (2020)
21. Lessmann, N., Van Ginneken, B., De Jong, P.A., Išgum, I.: Iterative fully convolutional neural networks for automatic vertebra segmentation and identification. Med. Image Anal. **53**, 142–155 (2019)
22. Payer, C., Stern, D., Bischof, H., Urschler, M.: Coarse to fine vertebrae localization and segmentation with spatial configuration-net and u-net. In: VISIGRAPP (5: VISAPP), pp. 124–133 (2020)