

Philipp Ernst^{1,2}, Georg Hille¹, Christian Hansen¹, Klaus Tönnies¹, Marko Rak¹

A CNN-Based Framework for Statistical Assessment of Spinal Shape and Curvature in Whole-Body MRI Images of Large Populations

Preprint Version

¹Department of Simulation and Graphics, University of Magdeburg, Germany

²Institute of Technical and Business Information Systems, University of
Magdeburg, Germany

A CNN-Based Framework for Statistical Assessment of Spinal Shape and Curvature in Whole-Body MRI Images of Large Populations

Philipp Ernst^{1,2}, Georg Hille¹, Christian Hansen¹, Klaus Tönnies¹, Marko Rak¹

¹Department of Simulation and Graphics, University of Magdeburg, Germany

²Institute of Technical and Business Information Systems, University of Magdeburg, Germany

Abstract. The extraction of spines from medical records in a fast yet accurate way is a challenging task, especially for large data sets. Addressing this issue, we present a framework based on convolutional neural networks for the reconstruction of the spinal shape and curvature, making statistical assessments feasible on epidemiological scale. Our method uses a two-step strategy. First, anchor vertebrae and the spinal centerline in between them get extracted. Second, the centerlines are transformed into a common coordinate system to enable comparisons and statistical assessments across subjects. Our networks were trained on 103 subjects, where we achieved accuracies of 3.3mm on average, taking at most 1s per record, which eases the handling of even very large cohorts. Without any further training, we validated our model on study data of about 3400 subjects with only 10 cases of failure, which demonstrates the robustness of our method with respect to the natural variability in spinal shape and curvature. A thorough statistical analysis of the results underpins the importance of our work. Specifically, we show that the spinal curvature is significantly influenced by the body mass index of a subject. Moreover, we show that the same findings arise when Cobb angles are considered instead of direct curvature measures. To this end, we propose a generalization of classical Cobb angles that can be evaluated algorithmically and can also serve as a useful (visual) tool for physicians in everyday clinical practice.

1 Introduction

Recently, medical images have become an integral part of public health studies to get a broader insight into the data [12, 13]. However, they are not suitable for direct analyses as the important information is not obtainable directly. For this reason, frameworks need to be established that extract information suitable for epidemiological purposes. Furthermore, non-experts in image processing should be able to apply the methodology. Our goal is to provide such a framework for quantification of spinal curvatures on large scale data sets. Following the idea behind visual analytics [8], the framework shall be suitable for the exploration of data sets with respect to certain factors to derive hypotheses on potential

relations and to later validate these using statistical tests. We will show that our framework is well-suited to implement this workflow (exploration \rightarrow hypothesis \rightarrow testing). To this end, we will investigate the relation between the body mass index (BMI) and spinal curvature. Due to the heterogeneity that can be expected from public health studies, our framework is based on convolutional neural networks (CNNs), which achieve state-of-the-art results for many medical image processing tasks [5] with large natural variability between subjects.

Related Work Many attempts have been made in the scope of spine analysis. Han et al. [2] use a Recurrent GAN for simultaneous semantic segmentation of spinal structures and classification of spinal diseases for 2D MR images. Hille et al. [3] present a hybrid level-set-based approach for robust and precise segmentation of vertebral bodies in 3D clinical routine MRI with minimal user interaction. Korez et al. [4] couple deformable models with CNNs for supervised segmentation of vertebral bodies from 3D MR spine images. Castro-Mateos et al. [1] introduce Statistical Interspace Models as an extension to Statistical Shape Models that take relative positions and orientations among objects into account and apply them to spine segmentation of CT images. Rak et al. [9] apply a combination of CNNs and graph cuts with star-convexity constraints to 3D patches of MR vertebra images to segment the whole spine. For further works please see the survey of Rak and Tönnies [10].

2 Method

Pretest Data Set For the development of our framework, a pretest data set of the Study of Health in Pomerania (SHIP) [13] was used. It includes T_1 -weighted whole spine MR images of 103 subjects that were acquired sagittally on a 1.5 T Siemens scanner with a field-of-view (FOV) of 50×50 cm and a voxel spacing of $1.1 \times 1.1 \times 4.4$ mm. We processed the images as follows: z-score normalization was applied to better generalize and the in-plane resolution was halved to $2.2 \times 2.2 \times 4.4$ mm to speed up training. Ground truth annotations for each subject were created by the authors. Specifically, the centers of all vertebrae from the topmost cervical vertebra (C1) to the first sacral vertebra (S1) were defined for each subject, whereby an isotropically resliced volume with a resolution of $1.1 \times 1.1 \times 1.1$ mm was used during creation to maximize ground truth precision.

Outline The first step of our method extracts a centerline probability map (CPM) of the spine using a V-Net-like CNN (Fig. 1) [7]. It is not favorable to directly extract the individual vertebrae via regression since their number may vary between subjects, which is difficult to reflect at the output of a neural network. As an example, this natural variability occurs in 5% of the Chinese population [14] and thus cannot be neglected in the context of public health studies. In parallel, a second CNN (Fig. 1 but with 2 output channels) extracts the positions of the anchor vertebrae (AVs), which in our case are C1 and S1.

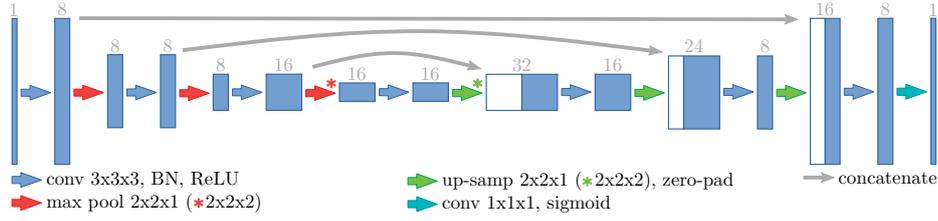


Fig. 1. Architecture of our CNN for extracting the centerline probability map. Blue boxes depict multi-channel feature maps. A white box corresponds to a copied feature map. The number above each box denotes the number of channels.

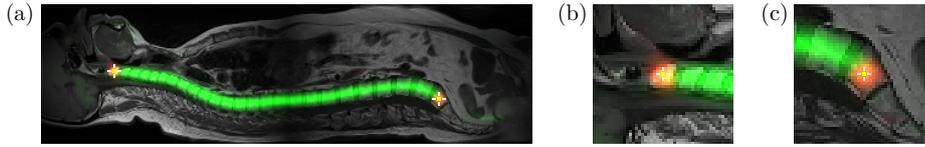


Fig. 2. (a) Predictions of our CNNs. Anchor vertebra CNN (red channel). Extracted anchor vertebra locations (crosses). Centerline probability map CNN (green channel). (b-c) Zoomed patches centered around the upper and lower vertebra, respectively.

These AVs can be found reliably due to their specific appearance and shape, which will be shown in the experiments. Given the trained CNNs, the centerline is extracted by following the CPM in between the found AVs. Afterwards, the reconstructed centerline is transformed into a reference system. This high-level spine representation enables simple visual as well as statistical comparability, as will be shown later on.

Anchor Vertebra Localization For localization of the AVs, the CNN is trained to predict one probability map for each AV from the whole spine MRI. To create these maps, a white voxel is inserted into a zero-volume at the position of the respective AV and a Gaussian filter ($\sigma = 6$ mm) is applied to yield a smooth blob around the AV location. The size of the kernel was chosen to roughly reflect the size of an average vertebra. After prediction, the AV center locations can be extracted easily by searching for the global maximum per channel (Fig. 2, red channel, crosses). Preliminary tests showed that this fuzzy segmentation task results in more accurate and robust localizations compared to a coordinate regression task, which would have been the intuitive strategy.

Centerline Calculation To create the CPM, we fit a cubic spline through the ground truth vertebra locations and rasterize the resulting curve into a volume that is initially filled with zeros. Afterwards, a Gaussian filter ($\sigma = 6$ mm) is applied to create a smooth ridge along the spine, which is then to be predicted via a dedicated CNN (Fig. 2 green channel). As for AV localization, this fuzzy



Fig. 3. Sagittal (left) and coronal projections (right) of the normalized centerlines of the about 3400 subjects of the SHIP study. The visualization encodes the distribution of shapes in boxplot-manner using different gray values. From white to black: median centerline, centerline range from 1st to 3rd quartile, centerline ranges for the 5th and 95th percentile. Dashed yellow lines separate the cervical, thoracic and lumbar area.

segmentation task yields more accurate and robust predictions compared to direct coordinate regression. After prediction, the actual centerline is defined by locations of maximal probability for each axial slice between the AVs. A shortest path algorithm is reasonable, too, and also more robust in case of ambiguous CPM predictions, which was not the case for any of our data.

Centerline Normalization For further analyses, the extracted centerlines are transformed into a common reference system by applying a similarity transformation, i.e. translation, rotation and isotropic scaling, to each centerline. The transformation maps the extracted AV locations onto fixed targets in the common reference system (Fig. 3), which can be implemented easily using Rodrigues’ rotation formula, yielding an angle- and ratio-preserving mapping by construction, which is suitable for later statistical analysis.

Cobb Angle Transformation To compare our framework with earlier work, we utilize Cobb angles, which are typically measured between inflection points. Based on preliminary tests, we found that the localization of these points and the subsequent angle measurement is error-prone when directly applied inside a neural network. For this reason, we generalized the idea of Zukić et al. [15]: For each pair of positions along the centerline, the signed angle between the direction vectors along the centerline is calculated (Fig. 4 (a)). We arrange the so-obtained values into a matrix representation (Fig. 4 (b)), the so-called Cobb angle transformation (CAT). It contains the (signed) Cobb angles between any pair of points along the spine and can be seen as a generalization of classic Cobb angles which are calculated between inflection points only. Using the CAT, we can also reconstruct classic Cobb angles by finding the maximum (unsigned) angle in the matrix representation.

Data Augmentation To cope with the rather small number of training samples available, a data augmentation without interpolation was applied in each epoch, consisting of random flips along the sagittal or transverse axis as well as random pixel-wise translations along the coronal or transverse axis. To constrain the translation during augmentation, we require that all vertebra locations remain inside the input domain even after augmentation.

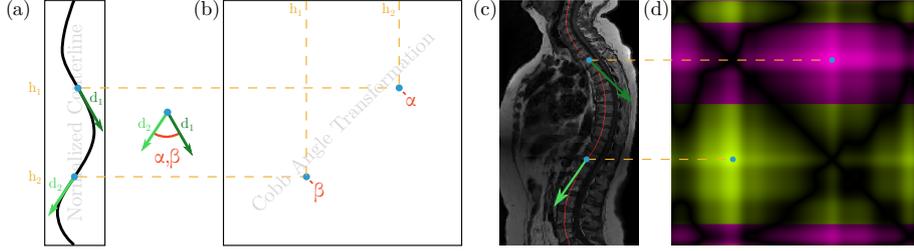


Fig. 4. (a) Normalized direction vectors d_1 and d_2 along the normalized centerline at positions h_1 and h_2 . (b) Cobb angle transformation contains $\alpha = \angle(d_1, d_2)$ and $\beta = \angle(d_2, d_1)$ at positions (h_1, h_2) and (h_2, h_1) where $\angle(x, y)$ defines the signed angle between x and y . (c-d) Example for (a-b) using $\angle(a, b) = \text{sgn}(a_x) \cdot \arccos(\langle a, b \rangle)$ depicting the part of the vertebral column with largest Cobb angle with respect to the sagittal plane. Cobb angles in the coronal plane are defined analogously.

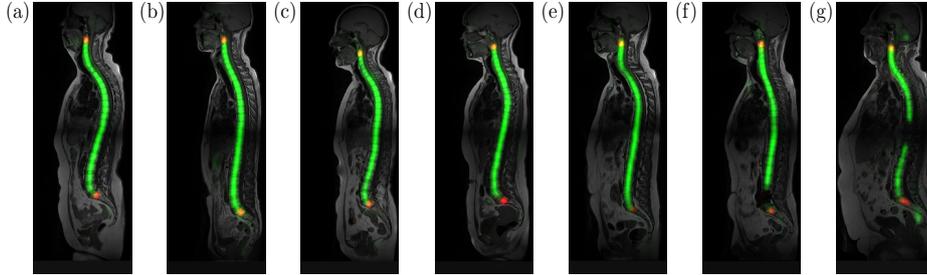


Fig. 5. Predictions on unseen data from the main SHIP study. Almost all results were classified as correct (a-d); few cases suffered from, for example, a weak prediction of the lower anchor vertebra (e), disappearing vertebrae due to acquisition artifacts (f) or bad illumination in the thoracic area (g).

Pretest Data Results After training both CNNs using stochastic gradient descent (learning rate = 0.01, Nesterov momentum = 0.7) [11] on the fuzzy Dice loss [7] for 600 epochs (AV CNN) and 200 epochs (CPM CNN) on the training data (82 of 103 samples), our method achieves a mean absolute error of 2.7 mm with respect to the ground truth on the validation data (21 of 103 samples), which is already well below the inter-slice spacing for our data and thus sufficient for our needs.

3 Experiments

Study Data Sets After training both networks on the pretest data, the main SHIP data is evaluated without any further training. It consists of about 3400 records which also include interviews on the health status and laboratory data. The main SHIP data is stratified based on geographical regions, subject age and

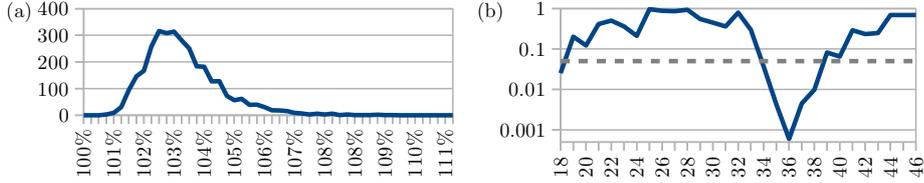


Fig. 6. (a) Histogram of lengths of normalized centerlines in the sagittal plane (100 % being a perfectly straight spine). (b) p-values of Mann-Whitney U tests on BMI thresholds, comparing the normalized centerline lengths above and below each BMI threshold for statistical differences; significance level $\alpha = 0.05$ (gray, dashed).

sex, which leads to near-uniform distributions across these attributes. Regarding subject weights, the distribution is nearly Gaussian.

Study Data Results Since the main SHIP data set does not include any ground truth, we assessed the result quality (Fig. 5) by visual inspection. To this end, the computed centerlines were used to create sagittal and coronal curved planar reformations yielding only two slices per subject with centerline overlay. A result was classified as success if the centerline stayed inside the vertebral column and its appearance seemed natural (e.g. no visual discontinuities). Only 10 of the about 3400 records (0.3 %) could not be classified as success, which underpins the robustness of our method on unseen data and renders it applicable for large data sets. These failure cases are related to artifacts (Fig. 5 (e)) or bad illumination (Fig. 5 (f)).

Exploration and Hypotheses To verify the suitability of our method for epidemiological analysis workflows, we used our centerline normalization to derive reasonable hypotheses about the data. Having a look at the distribution of spines with respect to attributes like subject size and weight, we came up with the hypothesis that subjects with a predisposition to obesity tend to have more bent spines than non-obese subjects. To verify this exemplary hypothesis, we set up according statistical tests, the results of which will now be discussed.

Statistical Testing In our first experiment, we compare different body mass index (BMI) levels by spine curvature directly, exploiting our centerline normalization approach. In the common coordinate system, the centerline length directly corresponds to curvature. For statistical evaluation, the Mann-Whitney U (MWU) test [6] was used which tests whether one random variable is statistically larger than the other one. Specifically, we test whether subjects with smaller BMIs have straighter spines. To this end, we split the set of subjects at each BMI level and perform MWU tests for each split. In Fig. 6 (right), the resulting p-values are depicted graphically. We observe significant results ($\alpha = 0.05$) for $34 \leq \text{BMI} \leq 38$. Looking at the actual centerline lengths, we observe that subjects of obesity class 1 or lower have significantly straighter spines

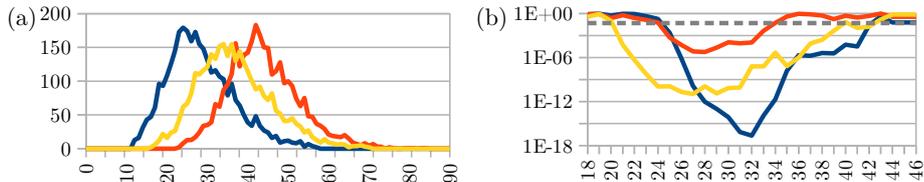


Fig. 7. (a) Histogram of maximum of the Cobb angle transformation in degrees. (b) p-values of Mann-Whitney U test on BMI thresholds, comparing the Cobb angles above and below each BMI threshold for statistical differences in the cervical (blue), thoracic (red) and lumbar (yellow) area; significance level $\alpha = 0.05$ (gray, dashed).

(\varnothing 102.5% length) than those of obesity class 2 or higher (\varnothing 103.2% length). This finding looks intuitive at first glance, but due to the lack of large scale assessments, we have not found any consolidation in literature. This underpins the need for frameworks like ours. We set up another experiment based on the CAT for further underpinning. Here, the Cobb angles were evaluated for the cervical, thoracic and lumbar area, using the very same procedure as described above. From the results (Fig. 7), we observe significant differences when splitting subjects at BMI levels $\{25 - 42\}$, $\{25 - 33\}$, $\{21 - 42\}$ for the three respective sections (Fig. 7 (b)). Comparing the estimated Cobb angles after the subjects split, the cervical (less obese: 27° vs. more obese: 23°) and the thoracic area (less obese: 43° vs. more obese: 41°) are bent less and the lumbar area is bent more (less obese: 33° vs. more obese: 35°) for subjects with higher obesity levels.

4 Conclusion

Our goal was to enable analysis of large image data sets in the context of epidemiological public health studies. Specifically, we provide a CNN-based framework to extract the whole spinal centerline from MR images in a fast yet accurate way. To ease the comparison of the extracted centerlines, we introduced a normalization step, which maps all centerlines into a common reference system. The latter can be used to derive hypotheses about the data, which are then tested statistically. Furthermore, we generalized standard Cobb angles into the so-called Cobb angle transformation, which proves a useful tool for detecting, classifying and quantifying spinal malformation at a glance in clinical daily routine. To underpin the importance of frameworks like ours, we derived hypotheses about the spine curvature on an epidemiological data set of about 3400 subjects. Using statistical tests, we found that the body mass index has a significant influence on the spinal curvature, i.e. subjects with at most obesity class 1 have straighter spines than more obese subjects. We also found that the spine in the cervical and thoracic section is bent less the higher the body mass index gets. Further tests, which were beyond the scope of this work, showed that even normal subjects have small but significant scoliotic tendencies (Cobb angles of -7° to 9°) and spines of females are bent more than those of males. Of course, with more

attributes becoming available in the course of a public health study, more interesting questions will arise. For example, whether the curvature of the spine is influenced by diseases not related to the spine or by certain human habits.

Acknowledgments We thank all parties and participants of the Study of Health in Pomerania. This work is funded by the EU and the federal state of Saxony-Anhalt, Germany (ZS/2016/08/80388 and ZS/2016/04/78123) as part of the initiative ‘Sachsen-Anhalt WISSENSCHAFT Schwerpunkte’. This work was conducted within the context of the International Graduate School MEMORIAL at Otto von Guericke University Magdeburg, Germany, kindly supported by the ESF under the program ‘Sachsen-Anhalt WISSENSCHAFT Internationalisierung’ (ZS/2016/08/80646). We thank the NVIDIA Corporation for donating the Titan Xp used for this research.

References

1. Castro-Mateos, I., Pozo, J.M., Pereañez, M., Lekadir, K., Lazary, A., Frangi, A.F.: Statistical interspace models (sims): Application to robust 3d spine segmentation. *IEEE Trans Med Imag* (2015)
2. Han, Z., Wei, B., Mercado, A., et al.: Spine-gan: Semantic segmentation of multiple spinal structures. *Med Imag Anal* (2018)
3. Hille, G., Saalfeld, S., Serowy, S., Tönnies, K.: Vertebral body segmentation in wide range clinical routine spine mri data. *Compt Methods Prog Biomed* (2018)
4. Korez, R., Likar, B., Pernuš, F., Vrtovec, T.: Model-based segmentation of vertebral bodies from mr images with 3d cnns. In: *MICCAI* (2016)
5. Litjens, G.J.S., Kooi, T., Bejnordi, B.E., et al.: A survey on deep learning in medical image analysis. *Med Imag Anal* (2017)
6. Mann, H.B., Whitney, D.R.: On a test of whether one of two random variables is stochastically larger than the other. *Annals Math Statis* (1947)
7. Milletari, F., Navab, N., Ahmadi, S.: V-net: Fully convolutional neural networks for volumetric medical image segmentation. *Int Conf 3D Vision* (2016)
8. Pak, C.W., Thomas, J.: *Visual Analytics*. *IEEE Compt Graph Appl* (2004)
9. Rak, M., Steffen, J., Meyer, A., et al.: Combining convolutional neural networks and star convex cuts for fast whole spine vertebra segmentation in mri. *Compt Methods Prog Biomed* (2019)
10. Rak, M., Tönnies, K.D.: On computerized methods for spine analysis in mri. *Inter J CARS* (2016)
11. Robbins, H., Monro, S.: A stochastic approximation method. *Annals Math Stat* (1951)
12. Tönnies, K.D., Gloger, O., Rak, M., et al.: Image analysis in epidemiological applications. *it* (2015)
13. Völzke, H.: Study of health in pomerania (ship). *Bundesgesundheitsblatt - Gesundheitsforschung - Gesundheitsschutz* (2012)
14. Yan, Y.Z., Li, Q.P., Wu, C.C., et al.: Rate of presence of 11 thoracic vertebrae and 6 lumbar vertebrae in asymptomatic Chinese adult volunteers. *J Orthopae Surg Res* (2018)
15. Zukić, D., Vlasák, A., Egger, J., et al.: Robust detection and segmentation for diagnosis of vertebral diseases using routine mr images. *Comput Graph Forum* (2014)