VERPEX: Anatomical Landmark Extraction on 3D Vertebrae exploiting Segmentation Masks

Hendrik Möller[1,2], Alissa Yuxuan Wang[1], Robert Graf[1,2], Kati Nispel[1], Matan Atad[1,2], Bjoern Menze[3], Daniel Rueckert[2,4], Jan S. Kirschke[1], and Tanja Lerchl[1]

[1] Dept. for Interventional and Diagnostic Neuroradiology, TUM University Hospital
[2] Chair for AI in Healthcare and Medicine, Technical University of Munich (TUM) and TUM University Hospital
[3] Department of Quantitative Biomedicine, University of Zurich
[4] Department of Computing, Imperial College London
hendrik.moeller@tum.de

Abstract. Digital twins of the human spine require accurate, subject-specific anatomical models to enable personalized biomedical simulation and clinical decision support. A critical prerequisite is the precise extraction of 3D anatomical landmarks, which define spinal alignment, vertebral orientation, and attachment points for ligaments and muscles. This work presents a two-step approach for predicting 3D landmarks from clinical CT and vertebra segmentation masks. It employs a DenseNet as the backbone for feature extraction and initial predictions, which are further refined by a transformer. This transformer receives local patch features, initial predictions, global features from the DenseNet, and the vertebra label to predict offset vectors, utilizing both global and local context and explicitly learning the relationship between landmarks. Our results show that prior knowledge from surface segmentation annotations can be successfully exploited to improve landmark accuracy, achieving average L1 errors of close to 1 mm. This significantly outperforms the baselines H3DE-Net and PRUnet3D. This work contributes to the digital twin ecosystem by providing a robust, automated pipeline that enables mechanical spine simulation tailored to the subject's individual anatomy.

Keywords: Landmark Extraction · Biomechanical Modeling · Segmentation · Spine.

1 Introduction

Biomechanical simulations have significantly contributed to our understanding of spinal health [2,16]. Numerical models enable in-depth analysis of spinal loading under various conditions, providing valuable insights into the mechanisms behind back pain and degeneration [12,15,19]. The spine, as the body's central load-bearing structure, is exposed to various factors that influence load transfer and distribution, including spinal curvature, vertebral shapes, and muscle and

ligament attachments [1,18]. The anatomical landmarks representing these characteristics are highly individual, and their accurate inclusion in biomechanical models is crucial to reflect the complexity of clinical reality [13]. Even deviations of a few millimeters can lead to adverse mechanical effects by inducing unrealistic forces on joints due to incorrect lines of action of ligament segments.

While medical images such as CT or MR provide valuable information, manual annotation of landmarks is time-consuming, error-prone, and subject to personal bias, and it requires expert knowledge to ensure anatomical accuracy. As a consequence, the vast majority of published models are generic, and highly individualized models are only available in small sample sizes [12]. However, to draw meaningful conclusions from simulations, analyzing large cohorts while accounting for their inherent diversity is essential. Therefore, automated methods are needed to ensure robust, consistent, and individualized landmark extraction from medical imaging.

Deep learning is the dominant approach for anatomical landmark detection in medical imaging, commonly using heatmap-based or direct regression methods [24,25]. Heatmap-based approaches offer strong spatial regularization and robustness to annotation noise but are computationally expensive in volumetric settings and prone to discretization effects [10]. Direct regression methods are more efficient yet often lack explicit spatial constraints, reducing stability in complex anatomical regions [26]. CNNs effectively model local features but struggle with long-range dependencies [26]. Transformer-based models address this via self-attention and improved global context modeling [3], enabling explicit learning of anatomical relationships [10]. However, when used alone, they may lose fine-grained spatial precision and remain computationally demanding.

Inspired by Sumon et al. [22], we propose a hybrid two-stage coarse-to-fine framework that utilizes heatmap-based regression with transformer-based refinement for landmark estimation, thereby combining global context with spatial accuracy. More importantly, we show that replacing the input with segmentation annotations to guide the model improves precision. Finally, the code is publicly available here, and model weights are shared here (examples) and upon request, enabling landmark generation on external datasets.

Fig. 1. Example mesh visualization of a sample of the dataset showing the POIs as dark spheres in 3D. The different colors of the vertebral surface represent the different subregions available in the semantic annotation of each vertebra.

2 Methodology

2.1 Dataset

We utilized a private in-house dataset. Informed consent was waived by the local ethics committee (593/21 S-NP). It consists of 561 vertebrae across 36 subjects with whole-spine 3D CT images and semantic and instance segmentation annotations for the vertebrae, generated using the Bonescreen SpineR tool (Bonescreen GmbH, Munich, Germany) based on Sekuboyina et al. [21]. Semantic annotation includes vertebral subregions, i.e., distinguishes between the vertebral corpus, arch, and processes (see Fig. 1).

The dataset comprises multiple spine segments per subject, spanning T1-L5, with 35 points of interest (POIs) per vertebra representing specific landmarks for ligament and muscle attachments, as well as intervertebral joint definition [11]. To create the dataset labels, we annotated POIs using the deterministic approach from Graf et al. [5] and subsequently manually reviewed each sample in Slicer3D [17] (see Fig. 1). The quality of each landmark was assessed based on anatomical correctness and intra- and intervertebral consistency relative to the surrounding landmarks, ensuring smooth lines of action for the respective attached ligament/muscle. Of all points, ∼90% were considered good or great, leading to the inclusion of 18,819 landmarks. Notably, all POIs, by definition, are located on the surface of each vertebra with sub-voxel accuracy.

We trained on 442 vertebra-level samples and tested on 119 samples (subject-wise split), resulting in a total test set of 3,783 landmarks after exclusions.

For preprocessing, we resample all images to a 1 mm^3 isotropic space and reorient them consistently to (Left, Anterior, Superior) order. Then, we create 3D cutouts around each vertebra with spatial dimensions (128, 128, 144). The center of each cutout is the center of mass of the corresponding vertebra segmentation annotation. We extract the surface segmentation by applying a 1-voxel erosion to the full segmentation mask and subtracting the result. This surface annotation is binary and does not include subregion distinctions.

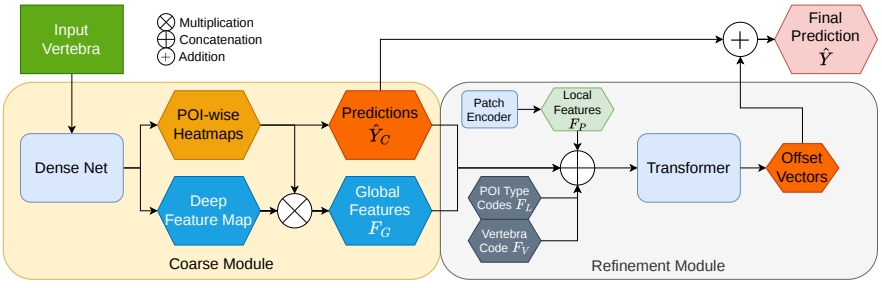

Fig. 2. Architecture of our two-stage VERPEX framework for landmark detection.

2.2 Training method

Our model architecture follows a two-stage design: a DenseNet-based feature extraction [8] is run on voxel maps to produce coarse predictions, followed by a transformer-based refinement module (see the architecture in Figure 2).

In detail, the first stage generates initial coarse coordinate estimates while preserving global spatial information via a hybrid heatmap localization and coordinate regression approach. We employ a modified DenseNet backbone, removing the final global pooling and replacing the classification head with a convolutional output layer. With b being the batch size, n the number of POIs, s the spatial shape of the image, and f the number of feature channels, this layer produces POI-specific heatmaps of shape (b, n, s) and a global feature map of shape (b, f, s). The heatmaps are normalized using a spatial softmax, and coarse POI coordinates $\hat{Y}_C$ are extracted via soft argmax. Global feature vectors F_G of shape (b, n, f) are obtained by weighting the shared feature map against each corresponding normalized heatmap and summing over the spatial dimensions, yielding a landmark-specific feature vector that captures the visual context at the predicted POI location. These global feature vectors F_G are passed to the refinement module together with the coarse predictions $\hat{Y}_C$.

In the second stage, the refinement module enhances the initial coordinate estimates by considering inter-landmark relationships and surrounding spatial context using a transformer architecture. Local patch features F_P are extracted around each coarse POI prediction $\hat{Y}_C$. Specifically, with a 1 mm isotropic voxel spacing, a cubic $16 \times 16 \times 16$-voxel patch is cropped from the input volume and centered on the coarse prediction. If the patch extends beyond the image boundaries, zero-padding is applied. The cropped patch is processed by a lightweight 3D CNN, producing a 64-dimensional feature vector F_P. These features are passed on to the transformer encoder along with the global features F_G, coordinate embeddings F_C (a learned linear projection of the coarse coordinates $\hat{Y}_C$), vertebral level embeddings F_V, and POI type embeddings F_L. The refinement module predicts a coordinate offset, which is added to the coarse estimate $\hat{Y}_C$ to obtain the final POI prediction $\hat{Y}$. We denote our proposed approach as *Vertebra Points-of-Interest Extraction* (VERPEX).

As data augmentation, we applied random flipping along the left/right axes and random affine transformations, including rotation, shearing, translation, and scaling. Since our labels are left/right-dependent, we ensured that the reference points and segmentations were relabeled appropriately during flipping (e.g., the left costal attachment point is relabeled as the right costal attachment point). If reference points were moved outside the field of view during transformation, their loss contribution was set to zero. We trained for 200 epochs on an Nvidia A40 with an early-stopping patience of 25 and a batch size of 8. We employed the Wing-Loss [4] as our loss function. Excluded landmarks were masked out in the loss calculation and omitted from evaluation metrics. For the exact training setup and hyperparameters, see here.

We used Python 3.10, PyTorch v2.9.1, and TPTBox v0.4.2. The Wilcoxon signed-rank test, computed with SciPy v1.15.3, was used to assess statistical significance, with $p < 0.05$ considered significant.

2.3 Experiments

As a benchmark, we compare our architecture against H3DE-Net [9], a hybrid CNN-Transformer framework that integrates volumetric attention mechanisms, and an anchor-based regression model built on a pruned residual 3D U-Net backbone (PRUnet3D) [7].

We conducted an ablation study to evaluate both the influence of the input data and the individual components of the proposed architecture.

First, we compared the impact of different input modalities by training the model using four input types: CT images (VERPEX-CT), segmentation masks (VERPEX-seg), CT images masked with instance segmentation (VERPEX-CT-seg), and surface-only segmentation masks (VERPEX-surface). To further assess the benefit of contextual information, we additionally run experiments using CT images and surface masks that included neighboring vertebrae in the input (VERPEX-CT-N and VERPEX-surface-N). In this setting, landmarks belonging to the neighboring vertebrae were incorporated with a reduced loss weight of 20% relative to the POIs of the primary vertebra.

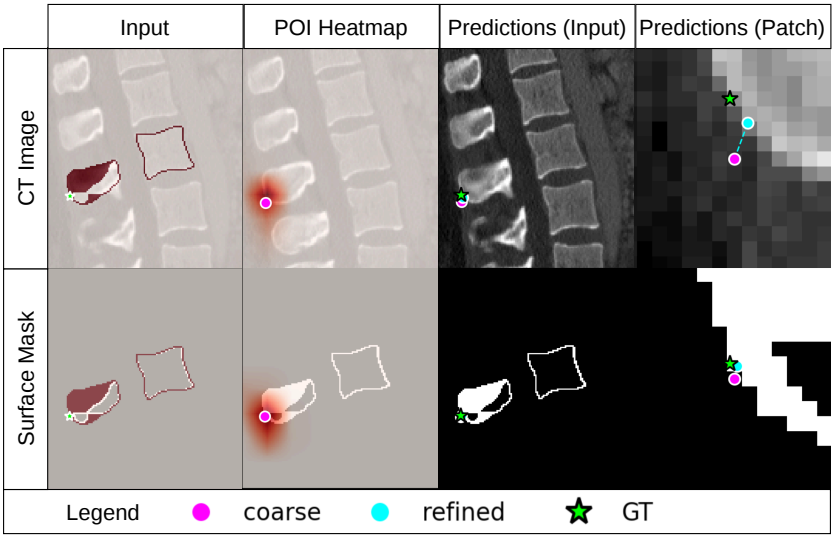

Fig. 3. Simplified 2D Visualization of how the VERPEX model outputs look in various stages on a sample and one specific landmark. The columns represent different input types (raw CT images, surface mask), while the rows, from left to right, show: input to the model, coarse heatmap, final prediction in the input space, and final prediction in the patch space.

To analyze the contributions of the individual components of the proposed model, we performed several architectural ablation experiments. Specifically, we evaluated the need for the refinement module by comparing coarse to refined predictions. Furthermore, we examined the importance of the different embeddings fed into the refinement module by selectively omitting them. We also assessed whether applying a surface projection (SP) from the predicted coordinates to the segmentation surface improves localization performance. Finally, we evaluated the impact of the loss function by comparing WingLoss to L1 and L2 losses.

Following standard practice, we evaluate the experiments by measuring the mean distance error (MDE) and the mean squared distance error (MSE). Additionally, we report the percentages of predicted points with error distances less than 2 mm ($P_{<2}$) and 4 mm ($P_{<4}$). To account for training variability, we train each setup five times with different seeds and additionally report the min and max MDE values (MDE/runs) across those five repeats. The other metrics are averaged across the five repeated training runs.

Since we excluded POIs from the reference data due to bad quality, we verified how our model performs on these samples and points. To this end, an experienced biomechanical engineer rated a randomly selected subsample of 50 POIs that were excluded from the reference data due to insufficient accuracy using a 5-point Likert scale (1 = poor accuracy, 5 = high accuracy). The same expert then rated the corresponding predictions from our best-performing model (based on validation metrics) using the same scale, and the two sets of ratings are compared.

3 Results

When trained on conventional raw CT images, the landmark baselines outperform VERPEX. However, when leveraging segmentation masks as input, VERPEX achieves significantly higher performance than all baselines at their best

Table 1. Comparison of different methods regarding the landmark detection on the test samples. Values are averaged across landmarks, samples, and across five training runs. Bold indicates best-performing values across the different methods. The arrow after a metric indicates whether higher or lower values are better. $P_{<2}$ and $P_{<4}$ are the percentage of points below 2 and 4 mm, respectively.

Setup	MDE ↓	MDE/runs	MSE ↓	$P_{<2}$ ↑	$P_{<4}$ ↑
H3DE-Net-CT	1.68 ± 0.96	$[1.66, 1.71]$	1.71	0.72	0.97
H3DE-Net-surface	1.76 ± 1.07	$[1.71, 1.78]$	1.80	0.69	0.96
PRUnet3D-CT	1.63 ± 0.98	$[1.58, 1.66]$	1.67	0.74	0.97
PRUnet3D-surface	1.60 ± 0.92	$[1.58, 1.64]$	1.64	0.75	0.97
VERPEX-CT	2.39 ± 2.85	$[1.91, 2.91]$	3.74	0.58	0.90
VERPEX-CT-seg	1.41 ± 0.83	$[1.32, 1.52]$	1.64	0.81	**0.99**
VERPEX-seg	1.36 ± 0.79	$[1.26, 1.41]$	1.57	0.84	**0.99**
VERPEX-surface	$\mathbf{1.28} \pm 0.74$	$[1.23, 1.31]$	**1.48**	**0.86**	**0.99**

configurations ($p < 0.001$), demonstrating that our architecture is specifically designed to exploit the additional structural information provided by segmentation (see Table 1). Qualitative examples are shown in Figure 3 and 4. Notably, neither H3DE-Net nor PRUnet3D benefits from surface segmentations, and in some cases, exhibits reduced performance. In contrast, training VERPEX on the segmentation surface (VERPEX-surface) of the target vertebra yielded the best overall results with an average MDE of 1.28. While 99% of predicted points had errors less than 4 mm, 86% had errors below 2 mm.

Our ablation experiments (see Table 2) show that incorporating neighboring vertebrae as additional context significantly reduced performance ($p < 0.001$). Furthermore, replacing WingLoss with traditional loss functions such as L1 or L2 also results in lower performance. Moreover, disabling transformer embedding

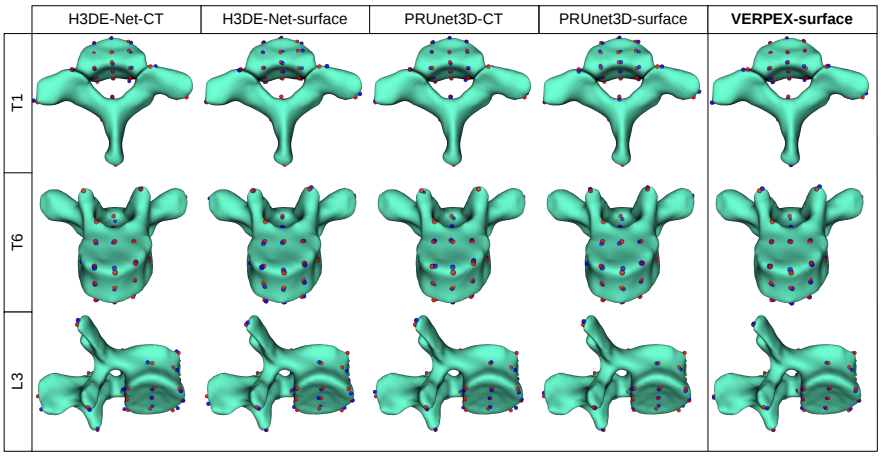

Fig. 4. Three random vertebrae (rows) from the test set and the corresponding baselines compared to our best setup (columns). Each sample shows the vertebra segmentation (cyan), reference annotation (red spheres), and the corresponding prediction (blue spheres). For each setup, we chose the model training with the best validation metrics for visualization.

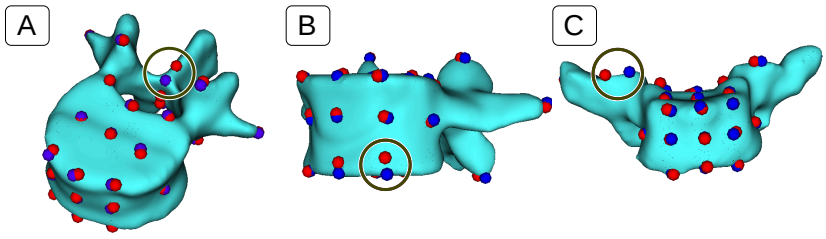

Fig. 5. Three random vertebrae A–C, the segmentation (cyan), reference annotation (red spheres), and our best prediction (blue spheres). Dark circles highlight POIs we excluded as incorrect.

Table 2. Ablation of our VERPEX-surface method, including architectural and training setup changes. Values are averaged across landmarks, samples, and across five training runs. The arrow after a metric indicates whether higher or lower values are better. $P_{<2}$ and $P_{<4}$ are the percentage of points below 2 and 4 mm, respectively. SP stands for Surface Projection.

Setup	MDE ↓	MDE/runs	MSE ↓	$P_{<2}$ ↑	$P_{<4}$ ↑
VERPEX-surface	1.28 ± 0.74	$[1.23, 1.31]$	1.48	0.86	0.99
VERPEX-surface-N	1.74 ± 1.15	$[1.59, 1.92]$	2.09	0.70	0.96
VERPEX-CT-seg-N	1.53 ± 1.03	$[1.44, 1.62]$	1.85	0.78	0.98
Only Coarse	1.41 ± 0.79	$[1.31, 1.45]$	2.61	0.82	0.99
Only Coarse with SP	1.33 ± 0.81	$[1.24, 1.37]$	2.41	0.83	0.99
Refined with SP	1.28 ± 0.76	$[1.23, 1.31]$	1.49	0.86	0.99
L1 Loss	1.41 ± 0.79	$[1.29, 1.52]$	1.62	0.82	0.99
L2 Loss	1.76 ± 0.93	$[1.53, 2.03]$	2.01	0.68	0.97
No Vertebra ID	1.47 ± 0.84	$[1.27, 1.87]$	1.56	0.79	0.98
No POI Embedding	1.57 ± 0.88	$[1.26, 1.74]$	1.82	0.74	0.98
No Local POI Features	1.43 ± 0.83	$[1.32, 1.52]$	1.65	0.81	0.99

inputs or relying solely on coarse predictions significantly decreases localization accuracy (all p-values < 0.001), thereby justifying the proposed refinement process. Finally, applying a post-processing surface projection (SP) to the predicted landmarks does not improve performance.

Regarding training time, with about 400 and 560 seconds per epoch, the H3DE-Net and PRUnet3D baselines were about five and seven times slower than our setup (∼81 sec per epoch), respectively. The average inference time for the logic-based deterministic approach used to create the initial reference annotations was ∼3.75 seconds. The baselines took about 0.27 seconds per vertebra, while our VERPEX-surface model required only ∼0.19 seconds on average (∼30% speed increase against the baselines, ∼95% increase against the deterministic approach).

Table 3. Results of the 5-point Likert scale rating of 50 random test cases for which the reference annotation was excluded.

Scale	(1) Bad	(2) Poor	(3) OK	(4) Good	(5) Great	Avg. ± SD
Reference	6	24	20	0	0	2.28 ± 0.66
Prediction	0	2	12	17	19	4.06 ± 0.88

Prediction accuracy for failed samples from the deterministic approach shows a strong increase (see Table 3). Expert 5-point Likert ratings for these samples in the reference data increased from an average of 2.28 ± 0.66 to 4.06 ± 0.88 (p-value < 0.001). For some examples, see Figure 5.

4 Discussion

Our study demonstrates the ability of VERPEX to predict anatomically relevant points of interest on the surfaces of vertebral bodies in CT images. It shows superior performance relative to the baselines, achieving an average error close to 1 mm. Furthermore, we demonstrate that segmentation masks help this model achieve much lower error than with traditional CT input alone.

Furthermore, qualitative analysis of the excluded reference points indicates that VERPEX not only accurately reproduces the original annotations but also performs reliably for POIs where the reference annotation failed. This improved consistency is particularly relevant for biomechanical modeling, as outliers can cause sudden, unrealistic changes in load transfer during simulations. For example, a $\sim$4 mm error in one of the two landmarks defining a vertical ligament segment across a single intervertebral joint can induce an artificial horizontal force component at the joint, thereby obscuring simulation results.

However, this work is not without limitations. The dataset lacks fractures or enumeration anomalies, limiting the model's applicability to real-world clinical data in its current form. Finally, our best-performing setup requires vertebra identification labels, which might not always be available. However, the ablation shows that even without this information, our system outperforms the baselines and yields significant results.

Future work includes establishing a public benchmark for vertebral POI detection using segmentation annotations, similar to VerSe [20], to enable standardized comparisons of methods. The predicted landmarks could support downstream tasks such as vertebral orientation estimation, rotation-invariant processing, and large-cohort normative analyses, with spatial outlier detection to flag anomalies (e.g., fractures, osteophytes, implants). Because vertebral segmentations can be derived from CT or MRI, the model may generalize across modalities if segmentation quality matches that of the training data. Although requiring an intermediate segmentation step, this improves robustness and accuracy. Future work may further enhance generalization by mimicking outputs of public segmentation tools (TotalSegmentator [23], VIBESegmentator [6], or SPINEPS [14]).

VERPEX demonstrates a robust, fast pipeline for the accurate extraction of spinal anatomical landmarks, enabling automated patient-specific biomechanical simulation and biometric analysis. Therefore, VERPEX enables large-cohort digital-twin studies to provide statistically relevant insights into spinal health.

References

1. Bruno, A.G., Anderson, D.E., D'Agostino, J., Bouxsein, M.L.: The effect of thoracic kyphosis and sagittal plane alignment on vertebral compressive loading. Journal of Bone and Mineral Research **27**(10), 2144–2151 (2012)
2. Bruno, A.G., Burkhart, K., Allaire, B., Anderson, D.E., Bouxsein, M.L.: Spinal loading patterns from biomechanical modeling explain the high incidence of vertebral fractures in the thoracolumbar region. Journal of Bone and Mineral Research **32**(6), 1282–1290 (2017)

3. Dosovitskiy, A., Beyer, L., Kolesnikov, A., Weissenborn, D., Zhai, X., Unterthiner, T., Dehghani, M., Minderer, M., Heigold, G., Gelly, S., Uszkoreit, J., Houlsby, N.: An image is worth 16x16 words: Transformers for image recognition at scale (2021), https://arxiv.org/abs/2010.11929
4. Feng, Z.H., Kittler, J., Awais, M., Huber, P., Wu, X.J.: Wing loss for robust facial landmark localisation with convolutional neural networks. In: Proceedings of the IEEE conference on computer vision and pattern recognition. pp. 2235–2245 (2018)
5. Graf, R., Lerchl, T., Nispel, K., Möller, H., Atad, M., McGinnis, J., Watrinet, J.M., Paetzold, J., Rueckert, D., Kirschke, J.S.: Rule-based key-point extraction for mr-guided biomechanical digital twins of the spine. In: International Workshop on Digital Twin for Healthcare. pp. 109–118. Springer (2025)
6. Graf, R., Platzek, P., Riedel, E.O., Ramschütz, C., Starck, S., Möller, H.K., Atad, M., Völzke, H., Bülow, R., Schmidt, C.O., et al.: Vibesegmentator: full body mri segmentation for the nako and uk biobank. European Radiology pp. 1–15 (2025)
7. He, T., Xu, G., Cui, L., Tang, W., Long, J., Guo, J.: Anchor ball regression model for large-scale 3d skull landmark detection. Neurocomputing **567**, 127051 (2024)
8. Huang, G., Liu, Z., van der Maaten, L., Weinberger, K.Q.: Densely connected convolutional networks (2018), https://arxiv.org/abs/1608.06993
9. Huang, Z., Tang, T., Xu, R., Wei, Y., Yang, W., Wang, S., Sun, X., Li, H., Yao, Q.: H3de-net: Efficient and accurate 3d landmark detection in medical imaging. In: 2025 IEEE International Conference on Bioinformatics and Biomedicine (BIBM). pp. 3710–3715. IEEE (2025)
10. Laitenberger, F., Scheuer, H.T., Scheuer, H.A., Lilienthal, E., You, S., Friedrich, R.E.: Cephalometric landmark detection using vision transformers with direct coordinate prediction. Journal of Cranio-Maxillofacial Surgery (2025)
11. Lerchl, T., El Husseini, M., Bayat, A., Sekuboyina, A., Hermann, L., Nispel, K., Baum, T., Löffler, M.T., Senner, V., Kirschke, J.S.: Validation of a patient-specific musculoskeletal model for lumbar load estimation generated by an automated pipeline from whole body ct. Frontiers in bioengineering and biotechnology **10**, 862804 (2022)
12. Lerchl, T., Nispel, K., Baum, T., Bodden, J., Senner, V., Kirschke, J.S.: Multibody models of the thoracolumbar spine: a review on applications, limitations, and challenges. Bioengineering **10**(2), 202 (2023)
13. Lerchl, T., Nispel, K., Bodden, J., Sekuboyina, A., El Husseini, M., Fritzsche, C., Senner, V., Kirschke, J.S.: Musculoskeletal spine modeling in large patient cohorts: how morphological individualization affects lumbar load estimation. Frontiers in Bioengineering and Biotechnology **12**, 1363081 (2024)
14. Möller, H., Graf, R., Schmitt, J., Keinert, B., Schön, H., Atad, M., Sekuboyina, A., Streckenbach, F., Kofler, F., Kroencke, T., et al.: Spineps—automatic whole spine segmentation of t2-weighted mr images using a two-phase approach to multi-class semantic and instance segmentation. European Radiology **35**(3), 1178–1189 (2025)
15. Nispel, K., Lerchl, T., Senner, V., Kirschke, J.S.: Recent advances in coupled mbs and fem models of the spine—a review. Bioengineering **10**(3), 315 (2023)
16. Oku, N., Demura, S., Tawara, D., Kato, S., Shinmura, K., Yokogawa, N., Yonezawa, N., Shimizu, T., Kitagawa, R., Handa, M., et al.: Biomechanical investigation of long spinal fusion models using three-dimensional finite element analysis. BMC Musculoskeletal Disorders **24**(1), 175 (2023)
17. Pieper, S., Halle, M., Kikinis, R.: 3d slicer. In: 2004 2nd IEEE international symposium on biomedical imaging: nano to macro (IEEE Cat No. 04EX821). pp. 632–635. IEEE (2004)

18. Putzer, M., Ehrlich, I., Rasmussen, J., Gebbeken, N., Dendorfer, S.: Sensitivity of lumbar spine loading to anatomical parameters. Journal of biomechanics **49**(6), 953–958 (2016)
19. Sciortino, V., Pasta, S., Ingrassia, T., Cerniglia, D.: On the finite element modeling of the lumbar spine: a schematic review. Applied Sciences **13**(2), 958 (2023)
20. Sekuboyina, A., Husseini, M.E., Bayat, A., Löffler, M., Liebl, H., Li, H., Tetteh, G., Kukačka, J., Payer, C., Štern, D., et al.: Verse: a vertebrae labelling and segmentation benchmark for multi-detector ct images. Medical image analysis **73**, 102166 (2021)
21. Sekuboyina, A., Rempfler, M., Kukačka, J., Tetteh, G., Valentinitsch, A., Kirschke, J.S., Menze, B.H.: Btrfly net: Vertebrae labelling with energy-based adversarial learning of local spine prior. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. pp. 649–657. Springer (2018)
22. Sumon, M.S.I., Islam, K.R., Hossain, S.A., Rafique, T., Ghosh, R., Hassan, G.S., Podder, K.K., Barhom, N., Tamimi, F., Chowdhury, M.E.: Self-cephalonet: a two-stage novel framework using operational neural network for cephalometric analysis. Neural Computing and Applications pp. 1–29 (2025)
23. Wasserthal, J., Breit, H.C., Meyer, M.T., Pradella, M., Hinck, D., Sauter, A.W., Heye, T., Boll, D.T., Cyriac, J., Yang, S., et al.: Totalsegmentator: robust segmentation of 104 anatomic structures in ct images. Radiology: Artificial Intelligence **5**(5), e230024 (2023)
24. Watchareeruetai, U., Sommana, B., Jain, S., Noinongyao, P., Ganguly, A., Samacoits, A., Earp, S.W., Sritrakool, N.: Lotr: face landmark localization using localization transformer. IEEE Access **10**, 16530–16543 (2022)
25. Zheng, Y., Liu, D., Georgescu, B., Nguyen, H., Comaniciu, D.: 3d deep learning for efficient and robust landmark detection in volumetric data. In: International conference on medical image computing and computer-assisted intervention. pp. 565–572. Springer (2015)
26. Zhou, X., Huang, Z., Zhu, H., Yao, Q., Zhou, S.K.: Hybrid attention network: An efficient approach for anatomy-free landmark detection. arXiv preprint arXiv:2412.06499 (2024)