Tailoring Loss to Boost Vertebra Centroid Localization and Classification in Sagittal Spinal Radiographs

Cliodhna Gartland[1,2] CLIODHNA.GARTLAND@UCDCONNECT.IE
[1] *Insight Research Ireland Research Centre for Data Analytics, Ireland*
[2] *University College Dublin, Ireland*

Fjorda Koromani[3,4,5] F.KOROMANI@ERASMUSMC.NL
[3] *Department of Internal Medicine, Erasmus MC, Rotterdam, The Netherlands*
[4] *Department of Epidemiology, Erasmus MC, Rotterdam, The Netherlands*
[5] *Department of Radiology and Nuclear Medicine, Erasmus MC, Rotterdam, The Netherlands*

John Healy[1,2] JOHN.HEALY@UCD.IE
Gennady Roshchupkin[4,5] G.ROSHCHUPKIN@ERASMUSMC.NL
Fernando Rivadeneira[3,4] F.RIVADENEIRA@ERASMUSMC.NL
Stephen J. Redmond[1,2] STEPHEN.REDMOND@UCD.IE

Editors: Under Review for MIDL 2025

Abstract

Osteoporotic vertebral fractures (OVFs) increase the risk of future fractures, morbidity, and mortality. However, manual interpretation of spinal radiographs is time-consuming and challenging. To address this, we propose an automated tool for vertebral centroid localization and classification in thoracolumbar sagittal spinal radiographs with varying fields-of-view, automating the initial input required for currently available semi-automated diagnostic tools and enhancing the efficiency of OVF assessment. To guide our model in its learning, we tested four loss functions to encourage focus on the vertebral centroid locations and to tailor our model to deal with limitations in our dataset, such as unlabeled visible vertebrae, that will aid in its generalizability. Our best performing model achieved a vertebral identification accuracy of 95.9% and a centroid localization RMSE for all correctly classified vertebrae of 17.08 pixels. Future work will leverage the outputs of this model for vertebral fracture detection in a fully-automated pipeline.

Keywords: Deep Learning, Loss, Spine, Radiographs

1. Introduction

Osteoporotic vertebral fractures (OVFs) are the most common type of osteoporotic fracture and indicate a higher risk of future osteoporotic fracture, morbidity, and mortality (Lentle et al., 2019). However, two-thirds of osteoporotic vertebral fractures are asymptomatic (Lenchik et al., 2004). Radiologists can have a significant impact on patient care through early detection of these asymptomatic fractures (Lenchik et al., 2004). From a radiograph, OVFs are currently detected through manual measurement of vertebral height loss, qualitative assessment, or use of semi-automated software that requires manual input or adjustment to analyze the radiograph (Lentle et al., 2019). Here we aim to develop the first step of an automated analysis process by localizing and classifying each vertebral centroid; future

work will examine the localized vertebrae for OVFs, either using existing semi-automated tools (Birch et al., 2015) or a fully automated pipeline for fracture diagnosis.

Currently, radiographs are the gold standard modality for OVF diagnosis (Capdevila-Reniu et al., 2021). Studies that localize and classify vertebrae in radiographs either used multiple models for the task (Cina et al., 2021) or rely on a specific field-of-view (FOV) within the input image, resulting in a model dependency on a particular vertebra always being visible from which to align the labels of all other identified vertebrae with (Kim et al., 2021, 2020; Sun et al., 2022; Fatima et al., 2022; Zou et al., 2023). The expectation of a certain FOV (i.e., containing a certain vertebra) within the input image will prevent these model from generalizing to other datasets. Instead, with this work, we aimed to: (1) develop a singular model that can both classify and localize the centroid of each vertebra for radiographs with varying FOV, and; (2) use dataset knowledge to tailor loss functions for enhanced performance and increased generalizability.

2. Methodology

For this work, we used 2,593 sagittal spinal radiographs from the ongoing Rotterdam Study (Hofman et al., 2015). Since two radiographs were often required to capture the full spine of a participant, the radiographs contain varying FOV and some overlap. Vertebrae centroid locations and labels were available for 13 vertebrae, from T4 to L4. However, each label was only made on one radiograph per participant, regardless of whether the vertebra was also visible on their other radiograph. More information on this dataset and our selection can be found in Appendix A. For all models, a training, validation, and test split of 80/10/10% was used. Each input image was resized to 264×264 pixels (px) and normalized using the training data. From the coordinates of each vertebral centroid, heatmaps were created to be used as targets for the models. These training targets (a 3D tensor) had dimensions equivalent to the dimensions of the input image (264×264 px) times a depth of 13 (i.e., one heatmap layer for each labeled vertebra). Each layer of this tensor consists of a Gaussian heatmap with a hotspot centered at the location of the vertebral centroid.

The U-Net architecture was selected as the baseline algorithm and was trained with four different loss functions, summarized in Appendix B: (1) UNet-Base served as the baseline for comparison with MSE loss; (2) UNet-L1 increased the model's focus on the regions directly surrounding each vertebral centroid; (3) UNet-L2 built on L1 by adding higher penalties for predicting centroids for vertebrae that were not visible within the input image and by not penalizing predictions for vertebra that were visible but unlabeled, and; (4) UNet-L3 built on L2 further by aiming to discourage predictions of the same centroid location for different vertebrae.

To assess the models using the test set, two key metrics were used: (1) the percentage of correctly classified vertebrae (vertebral ID accuracy) and; (2) the centroid location RMSE for all correctly classified vertebrae. To calculate the first metric, the centroid location for each heatmap layer was extracted and compared to the corresponding ground truth centroid location. If the difference was less than half the average distance between two neighboring ground truth centroids for that image, then the model was considered to have correctly classified the vertebra because the centroid prediction for that vertebra was within the boundary of the target vertebra with the same label. The second metric only uses

the centroids of correctly classified vertebra, as incorrectly classified vertebra could have predicted centroids in the center of a vertebra with a different ground truth label, but if assessed instead against the vertebra with the same label as the prediction, the error will be artificially inflated because it will be comparing with a centroid located on another vertebra.

3. Results

The results for each model are displayed in Table 1, with UNet-L2 as the best performing model. For context, the average distance between the ground-truth centroids of two consecutive vertebrae is 326±125 px (N=260), allowing us to conclude that our centroid predictions are in relatively-central locations for each vertebra. This can also be verified visually with the sample illustrations of results from UNet-L2 presented in Appendix C.

Table 1: Evaluation for each model calculated with a test set of 260 radiographs.

Model	Vertebra ID Accuracy	Centroid Location RMSE
UNet-Base	89.6 %	20.28 px
UNet-L1	94.7 %	17.17 px
UNet-L2	95.9 %	17.08 px
UNet-L3	95.2 %	19.65 px

4. Discussion and Conclusion

As expected, encouraging improved accuracy near the hotspots (i.e., vertebrae centroids) within each heatmap by tailoring the loss function, led to an improvement of at least 5% in the vertebral ID accuracy over the UNet-Base model. Further tailoring the loss function to improve robustness to how this dataset was labeled (i.e., with some visible vertebrae left unlabeled) proved beneficial; we noted that the dataset often contained vertebrae at the beginning or end of the labeled vertebrae set that had no labels for a particular image, and the proposed loss function accommodated missing labels in model output layers (L2 and L3). This encouraged the model to go beyond our labels and make predictions for vertebrae left unlabeled, which is how we would like the model to behave for all future images. Tailoring a loss function in this manner could be a valuable approach for other applications in which annotations are occasionally missing. However, when further tailoring the loss to avoid predicting centroid locations for two or more unique vertebrae on the same vertebra (L3), this led to a decrease in performance. This issue may be better addressed with basic post-processing heuristics.

Thus, we have successfully developed a single model for the task of detecting and classifying vertebral centroids in sagittal spinal radiographs with varying FOV and incorporated additional prior knowledge about our dataset into the loss function to further boost performance and increase generalizability as the model knows to label all vertebrae it detects instead of expecting the gaps in our dataset. Future work will focus on extracting region of interests (ROIs) based on our model's centroid predictions, to perform subsequent vertebra segmentation and fracture detection, providing medical staff with a fully-automated pipeline for efficient assessment of the spine.

Acknowledgments

This publication has emanated from research conducted with the financial support of Taighde Éireann - Research Ireland under grant numbers 17/FRL/4832 and SFI/12/RC/2289_P2. The Rotterdam Study is funded by Erasmus Medical Center and Erasmus University, The Netherlands Organization for the Health Research and Development (ZonMw), the Research Institute for Diseases in the Elderly (014-93-015, RIDE2), the Ministry of Education, Culture and Science, the Ministry for Health, Welfare and Sports, the European Commission (Directorate-General XII), and the Municipality of Rotterdam. For the purpose of Open Access, the author has applied a CC BY public copyright licence to any Author Accepted Manuscript version arising from this submission.

References

C. Birch, K. Knapp, S. Hopkins, S. Gallimore, and B. Rock. Spineanalyzer™ is an accurate and precise method of vertebral fracture detection and classification on dual-energy lateral vertebral assessment scans. *Radiography*, 21(3):278–281, 2015. doi: 10.1016/j.radi.2015.02.003.

A. Capdevila-Reniu, M. Navarro-López, and A. López-Soto. Osteoporotic vertebral fractures: A diagnostic challenge in the 21st century. *Revista Clínica Española*, 221:118–124, 2021. doi: 10.1016/j.rce.2019.09.006.

Andrea Cina, Tito Bassani, Matteo Panico, Andrea Luca, Youssef Masharawi, Marco Brayda-Bruno, and Fabio Galbusera. 2-step deep learning model for landmarks localization in spine radiographs. *Scientific Reports*, 11(1), 2021. doi: 10.1038/s41598-021-89102-w.

Joddat Fatima, Mashood Mohsan, Amina Jameel, Muhammad Usman Akram, and Adeel Muzaffar Syed. Vertebrae localization and spine segmentation on radiographic images for feature-based curvature classification for scoliosis. *Concurrency and Computation Practice and Experience*, 34(26), 2022. doi: 10.1002/cpe.7300.

Albert Hofman, Guy G. O. Brusselle, Sarwa Darwish Murad, Cornelia M. van Duijn, Oscar H. Franco, André Goedegebure, M. Arfan Ikram, Caroline C. W. Klaver, Tamar E. C. Nijsten, Robin P. Peeters, Bruno H. Ch. Stricker, Henning W. Tiemeier, André G, Uitterlinden, and Meike W. Vernooij. The rotterdam study: 2016 objectives and design update. *European Journal of Epidemiology*, 30:661–708, 2015. doi: 10.1007/s10654-015-0082-x.

Kang Cheol Kim, Hye Sun Yun, Sungjun Kim, and Jin Keun Seo. Automation of spine curve assessment in frontal radiographs using deep learning of vertebral-tilt vector. *IEEE Access*, 8:84618–84630, 2020. doi: 10.1109/ACCESS.2020.2992081.

Kang Cheol Kim, Hyun Cheol Cho, Tae Jun Jang, Jong Mun Choi, and Jin Keun Seo. Automatic detection and segmentation of lumbar vertebrae from x-ray images for compression fracture evaluation. *Computer Methods and Programs in Biomedicine*, 200, 2021. doi: 10.3389/fendo.2023.1132725.

Leon Lenchik, Lee F. Rogers, Pierre D. Delmas, and Harry K. Genant. Diagnosis of osteoporotic vertebral fractures: Importance of recognition and description by radiologists. *American Journal of Roentgenology*, 183(4):949–958, 2004. doi: 10.2214/ajr.183.4.1830949.

Brian Lentle, Fjorda Koromani, Jacques P. Brown, Ling Oei, Leanne Ward, David Goltzman, Fernando Rivadeneira, William D Leslie, Linda Probyn, Jerilynn Prior, Ian Hammond, Angela M. Cheung, and Edwin H. Oei. The radiology of osteoporotic vertebral fractures revisited. *Journal of Bone and Mineral Research*, 34(3):409–418, 2019. doi: 10.1002/jbmr.3669.

Yu Sun, Yaozhong Xing, Zian Zhao, Xianglong Meng, Gang Xu, and Yong Hai. Comparison of manual versus automated measurement of cobb angle in idiopathic scoliosis based on a deep learning keypoint detection technology. *European spine journal*, 31(8):1969–1978, 2022. doi: 10.1007/s00586-021-07025-6.

Lulin Zou, Lijun Guo, Rong Zhang, Lixin Ni, Zhenzuo Chen, Xiuchao He, and Jianhua Wang. Vltenet: A deep-learning-based vertebra localization and tilt estimation network for automatic cobb angle estimation. *IEEE Journal of Biomedical and Health Informatics*, 27(6):3002–3013, 2023. doi: 10.1109/JBHI.2023.3258361.

Appendix A. Dataset

For this work, we used a subset of data from the ongoing Rotterdam Study (Hofman et al., 2015). This study began in 1990, with participants all living within the Ommoord district in the city of Rotterdam, the Netherlands, and involves the collection of multiple health measurements from each participant every few years, including a sagittal spinal radiograph. The study has several cohorts, three of which we selected data. The first cohort (RSI) began in 1990 and includes 7,983 participants aged 55 years or older. The second (RSII) began in 2000 and includes 3,011 participants aged 55 years or older. The third (RSIII) began in 2006 and includes 3,932 participants aged 45-54 years. Scans were taken for each cohort every few years.

Co-author FK selected data from the first four scanning periods of RSI (RSI-1, RSI-2, RSI-3, RSI-4), the second scanning period of RSII (RSII-2), and the first scanning period of RSIII (RSIII-1). This selection was made by first identifying all scans that contained a vertebral fracture that had also already been labeled and then selecting a non-fractured scan from a different patient of the same sex and with a ± 1 year age difference within the same cohort and scan period; with this we have 1,312 scan selections. Participants usually have two radiographs per session, to capture vertebrae T4 to L4, giving us a total of 2,593 radiographs. These radiographs often overlap and have varying FOV. In addition, older radiographs are typically scanned copies of physical radiographic film, occasionally with markings, that have reduced image quality.

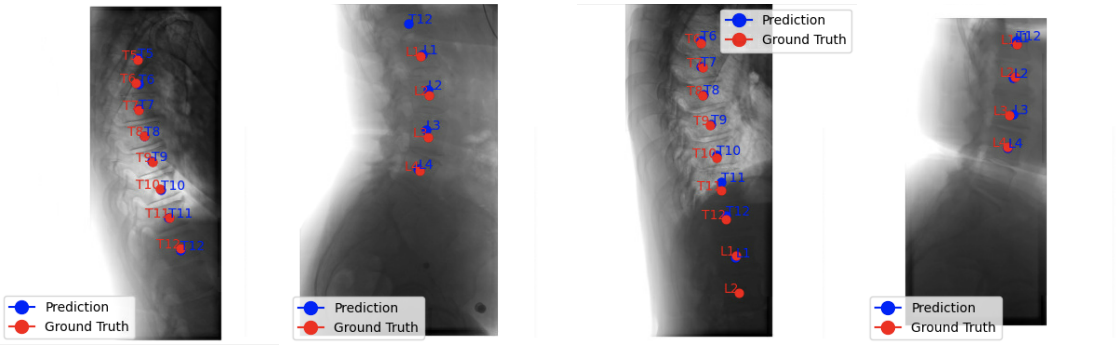
Vertebrae centroid locations and labels were also available within the study for vertebrae T4 to L4, totaling 13 vertebrae. However, these labels were only made on one radiograph per patient per scan period, regardless of whether the vertebra is also visible on the other radiograph of the rest of the spine.

Appendix B. Loss Functions

Table 2: Loss functions used for each trained model.

Loss	Loss Description	Formula	Purpose
MSE	Uniform weight for all pixel errors.	$f(y) = \frac{1}{N}\sum_{i=1}^{N}(y_i - \hat{y_i})^2$ where N represents the number of pixels in the 3D heatmap for an image, y_i represents the ground truth value, and $\hat{y_i}$ represents the predicted value.	Baseline for comparison.
L1	Pixels that had a ground truth value of 0.5 or greater (on a scale of 0.0 to 1.0) had their squared error weighted by a value of 10. Once the error is weighted, the mean is calculated.	$f(y) = \frac{1}{N}\sum_{i=1}^{N} X_i$ $X_i = \begin{cases}(y_i - \hat{y_i})^2 \times 10, & \text{if } y_i \geq 0.5\\ (y_i - \hat{y_i})^2, & \text{otherwise}\end{cases}$	Encourages focus on learning the areas directly around the centroids.
L2	L1 with the addition of all errors in layers where no vertebrae are visible weighted by a value of 5, and no loss calculation for layers before/after the first/last vertebra in a consecutively labeled segment.	$f(y) = \frac{1}{N}\sum_{j=1}^{13}\sum_{i=1}^{L_j} X_{i,j}$ $H_j = \sum_{i=1}^{L_j} y_{i,j}$ $X_{i,j} = \begin{cases}(y_{i,j} - \hat{y_{i,j}})^2 \times 10, & \text{if } y_{i,j} \geq 0.5\\ (y_{i,j} - \hat{y_{i,j}})^2 \times 5, & \text{if } H_j, H_{j+1}, H_{j-1} = 0\\ (y_{i,j} - \hat{y_{i,j}})^2 \times 0, & \text{if } H_j = 0, H_{j+1} \vee H_{j-1} \geq 0\\ (y_{i,j} - \hat{y_{i,j}})^2, & \text{otherwise}\end{cases}$ where L_j represents the pixels in a single heatmap layer j, i defines the current pixel within a heatmap layer, and H_j indicates whether the ground truth heatmap layer j contains a centroid or not.	Suppresses predictions of centroids in layers where there are no visible labeled vertebrae and prevents penalization of centroid predictions in visible but unlabeled vertebrae.
L3	L2 with all layers receiving a weighting of 10 in the location where pixels have a ground truth value of 0.5 or greater in any layer.	Same as L2 with one minor change: $X_{i,j} = \begin{cases}(y_{i,j} - \hat{y_{i,j}})^2 \times 10, & \text{if } y_{i,j} \geq 0.5 \textbf{ for any } j\\ (y_{i,j} - \hat{y_{i,j}})^2 \times 5, & \text{if } H_j, H_{j+1}, H_{j-1} = 0\\ (y_{i,j} - \hat{y_{i,j}})^2 \times 0, & \text{if } H_j = 0, H_{j+1} \vee H_{j-1} \geq 0\\ (y_{i,j} - \hat{y_{i,j}})^2, & \text{otherwise}\end{cases}$	Discourages predictions of the same centroid location for different vertebrae.

Appendix C. Sample Images



(a) (b) (c) (d)

Figure 1: Example predictions from best performing model UNet-L2 with: (a) correct predictions, (b) an extra correct prediction (T12), (c) a missing prediction (L2), and (d) two predictions (T12 and L1) on one vertebrae.