A vertebral compression fracture score based on deep generative contextual modeling

Michel Botros[1,2] M.BOTROS@AMSTERDAMUMC.NL
[1] *Department of Medical Imaging, Radboud University Medical Center Nijmegen, The Netherlands*
[2] *Department of Pathology, Amsterdam UMC, location Academic Medical Center, The Netherlands*
Matthieu Rutten[1,3] MATTHIEU.RUTTEN@RADBOUDUMC.NL
[3] *Department of Radiology, Jeroen Bosch Hospital, 's-Hertogenbosch, The Netherlands*
Twan van Laarhoven[4] T.VANLAARHOVEN@CS.RU.NL
[4] *Institute of Computing and Information Sciences, Radboud University, Nijmegen, The Netherlands*
Nikolas Lessmann[1] NIKOLAS.LESSMANN@RADBOUDUMC.NL

Editors: Under Review for MIDL 2022

Abstract

Grading of vertebral compression fractures most commonly relies on a semi-quantitative grading scale that defines fractures as height loss of the vertebral body. This paper presents an alternative approach that instead considers the three-dimensional shape of the vertebral body and expresses the abnormality of the shape on a scale from 0 to 100. The abnormality is expressed relative to the expected healthy shape, which is predicted by a deep generative model that is provided with contextual information, namely shape and orientation of surrounding vertebrae.

Keywords: spine, compression fractures, generative model, inpainting, anomaly detection

1. Introduction

Vertebral compression fractures are among the most common skeletal fractures caused by osteoporosis, a disorder characterized by a decrease in bone mineral density. These fractures may cause acute pain and may result in loss of function and mobility, but may also be asymptomatic. However, the presence of vertebral compression fractures is a strong predictor of future osteoporotic fractures, including fractures of other skeletal structures such as the hip (Klotzbuecher et al., 2000). To prevent this, osteoporosis management guidelines recommend pharmacological intervention when vertebral fractures are present.

Vertebral compression fractures are visible on various types of imaging exams, such as computed tomography (CT) and magnetic resonance imaging. The most commonly used classification for vertebral compression fractures is Genant's semi-quantitative grading scale (Genant et al., 1993), which assigns fracture grades based on the height loss of the vertebral body. In practice, the height loss is visually estimated or measured in sagittal images relative to neighboring normal vertebrae or prior imaging exams. A limitation of this method is the inherent subjectivity and the coarseness of the grading scale, in which normal vertebrae and mild and moderate fractures can differ by only a few percent of height loss.

This paper proposes an automatic method for computing a vertebral body abnormality score, intended as a more objective and more finely scaled alternative to Genant's grading scale. This score expresses the degree of abnormality of the three-dimensional shape of the vertebral body in comparison with the shape expected based on contextual information.

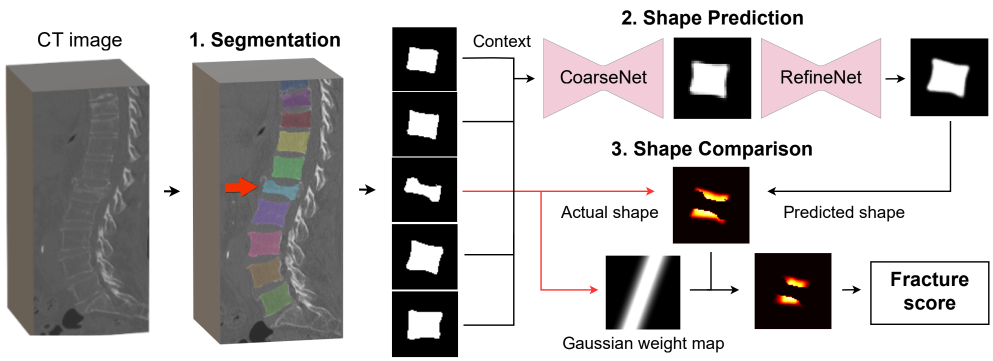

Figure 1: Overview of the proposed method for computing a vertebral abnormality score.

2. Methods

We use iterative instance segmentation (Lessmann et al., 2019) to segment and anatomically label the vertebral bodies in the image. Since the aim is to measure abnormality of the vertebral shape, subsequent steps operate on the segmentation masks instead of the images. Inspired by generative image inpainting (Yu et al., 2018), two generative models predict the shape and orientation of a vertebral body given shape and orientation of its neighbors. The first model is trained with a binary cross entropy loss to produce a coarse mask with approximately the expected position and orientation. The second model refines the coarse mask and is additionally trained with a Wasserstein GAN loss to enforce realistic shapes. Both models are constructed as encoder-decoder networks and are trained exclusively with unfractured vertebrae to introduce a bias toward generating only unfractured vertebrae. The segmentation masks of two neighboring vertebrae on either side of these normal vertebrae (or zeros) are stacked into a four-channel three-dimensional input for the coarse network. The same context plus the coarse prediction are provided as input to the refinement network.

The proposed abnormality score corresponds to the surplus volume of the predicted normal shape compared with the actual (potentially compressed) vertebral body. We suppress anteroposterior differences with a three-dimensional Gaussian weight mask rotated to the orientation of the vertebra. The final score is normalized with respect to the size of the vertebra and scaled to $[0, 100]$. The method is illustrated in Figure 1.

3. Results

We experimented with 115 CT images from the VerSe19 (Löffler et al., 2020) and xVertSeg (Ibragimov et al., 2017) datasets, excluding cervical scans. Ten percent of this data was used as internal test set, where we compared the abnormality score with Genant's grading scale (Figure 2). Additionally, 20 CT images from Radboudumc were used as external test set, where we compared the abnormality scores of fractured vertebrae (according to the radiology report) with those of neighboring normal vertebrae. For both test sets, average abnormality scores were significantly larger for fractured vertebrae (internal: 59.9 ± 24.8 vs. 10.1 ± 6.9; external: 75.1 ± 32.6 vs. 5.6 ± 7.9; both $p<0.001$). These initial results correspond well with the assessment by radiologists and underline the potential of quantitatively expressing abnormality of the vertebral body shape based on generative modeling.

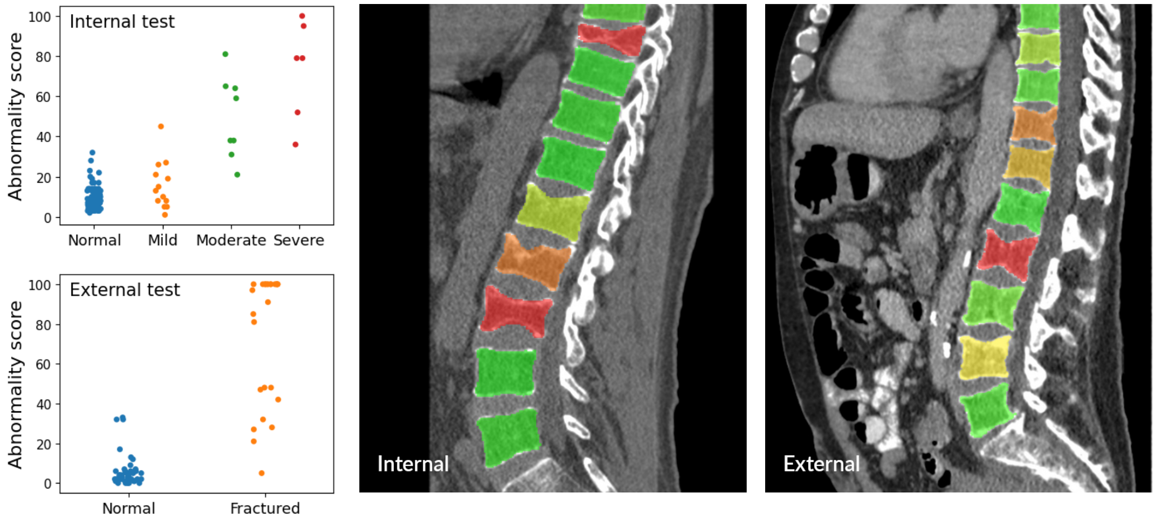

Figure 2: Scoring results for vertebrae in the internal and external test sets, according to fracture grades or fracture status (left). Colors in the examples (right) indicate scores from normal (0 = green) to abnormal (100 = red).

References

Harry K. Genant, Chun Y. Wu, Cornelis van Kuijk, and Michael C. Nevitt. Vertebral fracture assessment using a semiquantitative technique. *Journal of Bone and Mineral Research*, 8(9):1137–1148, 1993.

Bulat Ibragimov, Robert Korez, Bostjan Likar, Franjo Pernus, Lei Xing, and Tomaz Vrtovec. Segmentation of Pathological Structures by Landmark-Assisted Deformable Models. *IEEE Transactions on Medical Imaging*, 36(7):1457–1469, 2017.

Carolyn M. Klotzbuecher, Philip D. Ross, Pamela B. Landsman, Thomas A. Abbott, and Marc Berger. Patients with Prior Fractures Have an Increased Risk of Future Fractures: A Summary of the Literature and Statistical Synthesis. *Journal of Bone and Mineral Research*, 15(4):721–739, 2000.

Nikolas Lessmann, Bram van Ginneken, Pim A. de Jong, and Ivana Išgum. Iterative fully convolutional neural networks for automatic vertebra segmentation and identification. *Medical Image Analysis*, 53:142–155, 2019.

Maximilian T. Löffler, Anjany Sekuboyina, Alina Jacob, Anna-Lena Grau, Andreas Scharr, Malek El Husseini, Mareike Kallweit, Claus Zimmer, Thomas Baum, and Jan S. Kirschke. A Vertebral Segmentation Dataset with Fracture Grading. *Radiology: Artificial Intelligence*, 2(4):e190138, 2020.

Jiahui Yu, Zhe Lin, Jimei Yang, Xiaohui Shen, Xin Lu, and Thomas S. Huang. Generative Image Inpainting with Contextual Attention. In *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 5505–5514, 2018.