UDC 004.93'1

DEEP LEARNING BASED APPROACH FOR VERTEBRAE DETECTION ON SPINE X-RAY

R.M. Charapennikau
Student of the Department of Programming Technologies of the Belarusian State University
fpm.cherepen@bsu.by

R.M. Charapennikau

Student of Faculty of Applied Math and Computer Science, BSU. Machine Learning Engineer at BGSoft. Research interests: medical imaging applications of deep learning, generative models.

Abstract. In this paper two methods for vertebra detection on X-RAY images are presented. The key difference from standard object detection methods is that proposed ones (methods) are not limited to finding a rectangular bounding box, but can be used to detect an object with a bounding box of arbitrary quadrangle shape.

Our method achieves 0.988 Average Precision Score with 0.822 bounding box IoU.

Keywords: Deep learning, medical imaging, object detection, instance segmentation.

Introduction.

Statistics indicate that there are around 540 million adults who experience back pain of different levels of intensity [1].

Traditional methods of diagnosing diseases of the spine are radiography, computed tomography (CT) and magnetic resonance imaging (MRI). Despite the fact X-Ray is a total projection image and is not so accurate compared to CT and MRI, it is still the most popular way of primary diagnosis. This is due to the following facts: X-Ray is fast (the whole procedure can be done in several minutes), cheap and no special patient preparation is required.

Accurately detecting and identifying each of the individual vertebrae is essential for medical professionals in diagnosing and treating spinal conditions such as spinal fractures, herniated discs, scoliosis and spondylolisthesis. Traditionally, spine vertebra detection was done manually by radiologists, which is a time-consuming and labor-intensive process.

At the same time Neural Networks have proven to be a reliable way to solve many problems related to image processing.

In this work we research the possibility of applying them to solve the problem of automatic detection of vertebrae from spine X-Ray.

On top of the detection methods, a diagnosis decision making support system for medical professionals can be built. Decision support system aims to help to make the diagnostic process faster, more accurate, and less labor-intensive for medical professionals compared to manual diagnostics.

Related Work.

The object detection is the natural extension of object classification, which aims only at recognizing the object in the image. The goal of the object detection is to detect all instances of the predefined classes and provide its coarse localization in the image by axis-aligned boxes. The detector should be able to identify all instances of the object classes and draw a bounding box around it. It is generally seen as a supervised learning problem [2].

There are two major groups of object detection methods: Two-Stage Detectors (Faster-RCNN[3], Mask-RCNN[4]) and One-Stage Detectors (SSD[5], YOLO[6], RetinaNet[7]).

The key difference in these two types of object detection models is how they solve classification problem. Two stage models have separate region proposal subnetworks that generate bounding boxes, and later proposed regions are passed to classification subnetwork. At the same time one stage models solve region proposal and classification problems with a single subnetwork.

There also has been studies that explore automatic detection of spine diseases from X-Ray images. But they were limited to working in some particular regions of the spine, for example [8] is limited to Lumbar (S1, L1-L5 vertebrae, which is only 6 out of 32 vertebrae in the human body) spondylolisthesis.

In contrast, the goal of our research is to design a region-independent method of vertebrae detection, which can be further used to detect different pathologies.

Model evaluation.

Object detectors use multiple criteria to measure the performance of the detectors viz., frames per second (FPS), precision and recall. However, mean Average Precision (mAP) is the most common evaluation metric. Precision is derived from Intersection over Union (IoU), which is the ratio of the area of overlap and the area of union between the ground truth and the predicted bounding box. A threshold is set to determine if the detection is correct. If the IoU is more than the threshold, it is classified as True Positive while an IoU below it is classified as False Positive. If the model fails to detect an object present in the ground truth, it is termed as False Negative. Precision measures the percentage of correct predictions while the recall measures the correct predictions with respect to the ground truth.

$$Precision = \frac{True\ Positive}{True\ Positive + False\ Positive}$$

$$Recall = \frac{True\ Positive}{True\ Positive + False\ Negative}$$

Based on the above equation, average precision is computed separately for each class. To compare performance between the detectors, the mean of average precision of all classes, called mean average precision (mAP) is used, which acts as a single metric for final evaluation [2].

Vertebrae Detection.

As described previously, in most cases, when detecting, it is sufficient to use as bounding boxes rectangles, whose sides are parallel to the coordinate axes. However, in the problem under consideration, they are not accurate enough. For example, in the task of determining scoliosis and its degree it is necessary that the sides of the bounding box be directed in the same way as the vertebra is rotated. At the same time the upper and lower sides are not necessarily parallel to each other. Therefore, we need to use an arbitrary convex quadrangle as a bounding box.

We propose two ways of achieving this:

●Finding a rectangular bounding with a standard object detection model followed by a refinement step using one more model, that solves regression tasks on 4 key points coordinates (method 1).

●Using an instance segmentation model to build a segmentation mask for each vertebra and future approximation of this mask with a quadrangle bounding box (method 2).

Method 1

At the first step of this approach, the problem of detecting individual vertebrae in the image using a rectangular bounding box. For this we can use any of the standard architectures for object detection task, for example, FasterR-CNN, YOLO, SSD.

In the second step, each of the obtained bounding boxes is cut out from the original image, and used as input data for the neural network that determines the coordinates of 4 key points of each vertebra.

Best results in detection of individual vertebrae on X-Ray images was achieved by FasterR-CNN architecture with ResNet-50 backbone.

Experiments were also carried out using one-step methods, such as SSD, RetinaNet (more specifically, the experiments are described in Table 1). These methods, as expected, worked faster, but showed lower quality. Since in the problem under consideration the speed is not critical, preference was given to FasterR-CNN.

Table 1. Vertebrae detection experiments

Architecture	*Learning rate*	*Average Precision*	*Notes*
SSD (ResNet-50)	Starting: 3.0E-4 Decrease by 10 times every 5 epochs	Val: 0.921 Test: 0.918	ResNet-50 use pretrained ImageNet weights
RetinaNet(ResNet-50)	Starting: 3.0E-4 Decrease by 10 times every 5 epochs	Val: 0.948 Test: 0.949	ResNet-50 use pretrained ImageNet weights
Faster-RCNN (ResNet-50)	Starting: 3.0E-7 for backbone 3.0E-4 other layers Decrease by 10 times on plateau	**Val: 0.9872 Test: 0.9884**	ResNet-50 use custom pretrained weights

One of the considered pathologies is scoliosis. To calculate the degree of scoliosis it is necessary to calculate the angles along the boundaries of the bounding boxes. Thus, one of the important characteristics of the bounding box is the coincidence of the angles between the lines of the correct and the predicted box. Therefore, it is proposed during training to optimize not only the absolute distance between points (as is usually done in regression problems), but also the angles between the lines. It is proposed to use $AngeleLoss = sin^2(\theta)$ as a measure of the difference in angles, where θ is angle between predicted and correct bounding box boundary. And as a generalization of this loss function to the bounding box, the average value over 2 pairs of lines (predicted and correct bottom and top boundaries) is used.

Table 2. Vertebrae keypoints regression experiments

Architecture	*Loss function*	*IoU*
ResNet-101	Mean squared error	Val:0.783 Test:0.794
ResNet-101	Mean absolute error	Val:0.766 Test:0.776
ResNet-101	Mean squared error + Angle Loss	**Val:0.811 Test:0.822**
ResNet-101	Mean absolute error + Angle Loss	Val:0.786 Test:0.789

The results of applying method 1 to spine X-Rays are presented in Figure 1

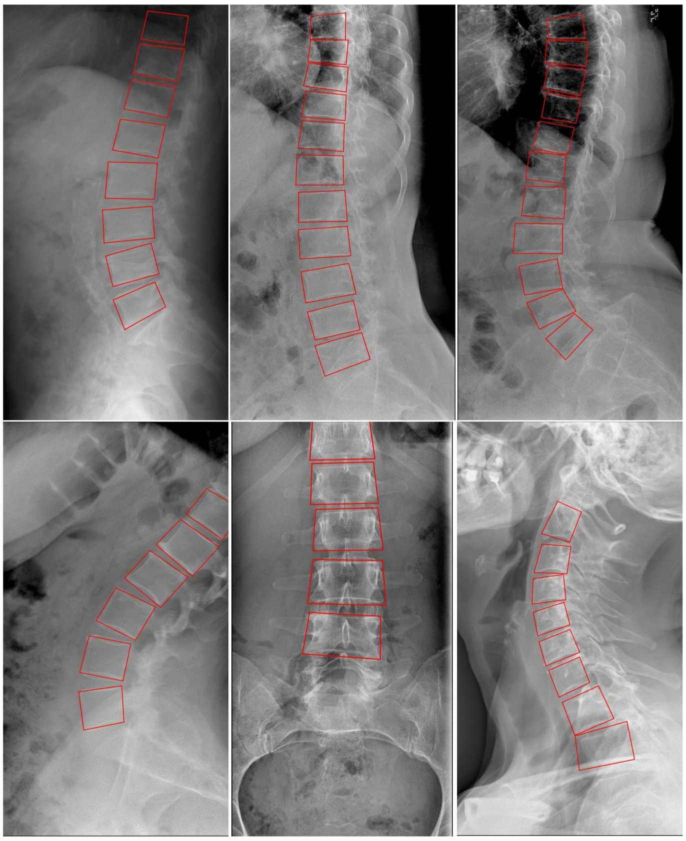

Figure 1. Results achieved with method 1

Method 2

The second approach for solving the problem under consideration is based on the instance segmentation. At the first step, masks are built for each of the vertebrae, and then they are approximated using quadrangles.

MaskR-CNN was used as an architecture for solving the instance segmentation problem. As a result, it was possible to achieve mask IoU **0.782** and bounding box Average Precision **0.984**.

To solve the problem of approximating masks using quadrangles, it is proposed to use gradient-free optimization methods.

The quadrangle is defined by 8 coordinates $(x_1, y_1, x_2, y_2, x_3, y_3, x_4, y_4)$, using these coordinates we can build a mask that is an approximation of the resulting mask. And as an objective function, we use the IoU value between the mask obtained using MaskR-CNN and the mask built using the coordinates of the quadrangle. The following methods were chosen for comparison: the Nelder-Mead method, differential evolution, and the simulated annealing method (the results of the comparison of approximation methods are shown in Table 3).

Table 3. Mask approximation experiments

Architecture	*Mean ± std execution time, s*	*Mean objective function value*
Nelder-Mead	0.0034 ± 0.008	0.925
Differential evolution	**0.1925 ± 0.0386**	**0.965**
Simulated annealing	0.2047 ± 0.0074	0.960

The results of applying method 2 are presented on Figure 2.

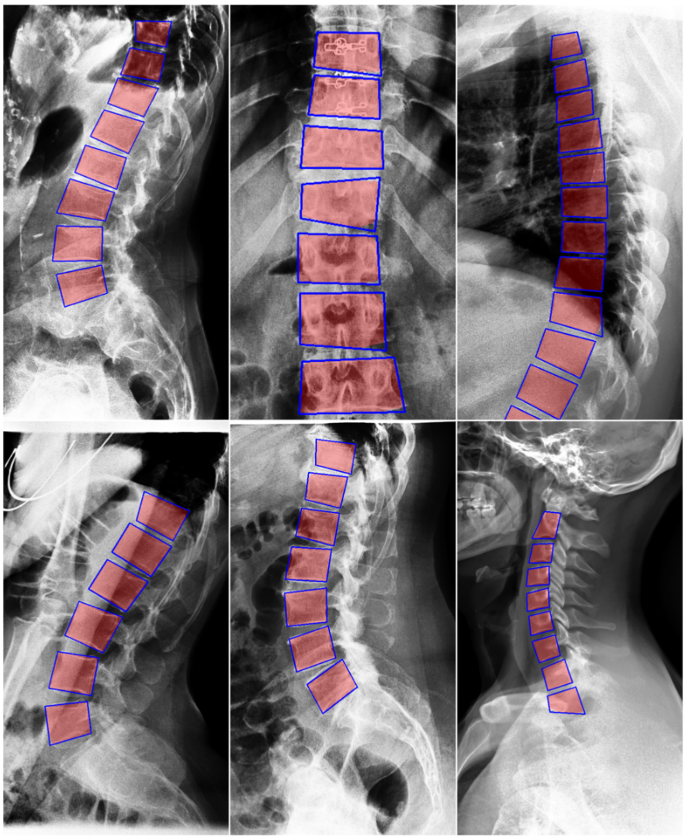

Figure 2. Results achieved with method 2

Conclusion.

In this research we have explored the possibility of using neural networks for the automatic detection of individual vertebrae from spine X-Ray images.

We proposed two methods for achieving this, one using a standard object detection model followed by a refinement step using another model that solves regression tasks on 4 keypoints coordinates, and the other using an instance segmentation model to build a segmentation mask for each vertebra and future approximation of this mask with a quadrangle bounding box.

Our experiments showed that the FasterR-CNN architecture with ResNet-50 backbone produced the best results in detecting individual vertebrae on X-Ray images (**0.988** Average Precision with **0.822** bounding box IoU).

The proposed methods can help medical professionals in diagnosing and treating spinal conditions such as spinal fractures, herniated discs, scoliosis, and spondylolisthesis.

Further research can be done to improve the accuracy of the proposed methods and extend it to detect different pathologies.

References

[1] 57 Back Pain Statistics: How Common Is Back Pain? [Electronic resource]. URL: https://www.crossrivertherapy.com/research/back-pain-statistics

[2] A Survey of Modern Deep Learning based Object Detection Models [Electronic resource]. URL: https://arxiv.org/abs/2104.11892

[3] Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks [Electronic resource]. URL: https://arxiv.org/abs/1506.01497

[4] Mask R-CNN [Electronic resource]. URL: https://arxiv.org/abs/1703.06870

[5] SSD: Single Shot MultiBox Detector [Electronic resource]. URL: https://arxiv.org/abs/1512.02325

[6] You Only Look Once: Unified, Real-Time Object Detection [Electronic resource]. URL: https://arxiv.org/abs/1506.02640

[7] Focal Loss for Dense Object Detection [Electronic resource]. URL: https://arxiv.org/abs/1708.02002

[8] Detection of Lumbar Spondylolisthesis from X-ray Images Using Deep Learning Network [Electronic resource]. URL: https://www.mdpi.com/2077-0383/11/18/5450

ОСНОВАННЫЙ НА ГЛУБОКОМ ОБУЧЕНИИ ПОДХОД К ОБНАРУЖЕНИЮ ПОЗВОНКОВ НА РЕНТГЕНОГРАММЕ ПОЗВОНОЧНИКА

Р.М. Черепенников
Студент кафедры
технологий
программирования БГУ

Кафедры технологий программирования
Факультет прикладной математики и информатики
Белорусскийо государственный университет.
Электронная почта: fpm.cherepen@bsu.by

Аннотация. В данной статье представлены два метода детекции позвонков на рентгеновских снимках. Основным отличием, предложенных методов, от существующих методов детекции является то, что они не ограничены поиском прямоугольной ограничивающей рамки, а решают задачу детекции с помощью ограничивающей рамки произвольной четырехугольной формы.

Предложенный метод достигает качества 0.988 метрики Average Precision при IoU ограничивающих рамок 0.822.

Ключевые слова: Глубокое обучение, методы медицинской визуализации, детекция объектов.