Bipartite Distance for Shape-Aware Landmark Detection in Spinal X-Ray Images

Abdullah-Al-Zubaer Imran[2,1], **Chao Huang**[1], **Hui Tang**[1], **Wei Fan**[1],
Kenneth M.C. Cheung[3], **Michael To**[3], **Zhen Qian**[1], **Demetri Terzopoulos**[2,4]

[1]Tencent Medical AI Lab, Palo Alto, CA, USA
[2]University of California, Los Angeles, CA, USA
[3]The University of Hong Kong, China
[4]VoxelCloud, Inc., Los Angeles, CA, USA

Abstract

Scoliosis is a congenital disease that causes lateral curvature in the spine. Its assessment relies on the identification and localization of vertebrae in spinal X-ray images, conventionally via tedious and time-consuming manual radiographic procedures that are prone to subjectivity and observational variability. Reliability can be improved through the automatic detection and localization of spinal landmarks. To guide a CNN in the learning of spinal shape while detecting landmarks in X-ray images, we propose a novel loss based on a bipartite distance (BPD) measure, and show that it consistently improves landmark detection performance.

1 Introduction

Scoliosis is an abnormal condition characterized by lateral spinal curvature. Early assessment and treatment planning is critical [4]. Conventionally, the assessment of scoliosis is performed manually by clinicians through the identification and localization of vertebral structures in spinal X-ray images. However, large inter-patient anatomical variation and poor image quality challenge clinicians to assess the severity of scoliosis accurately and reliably. Automated measurement promises to enable the reliable quantitative assessment of scoliosis.

Several spinal landmark detection methods are available in the literature: Conventional hand-crafted feature engineering [1] is a semi-automatic method involving several sub-tasks. Our approach is automatic convolutional neural network (CNN) models. The CNN model of Wu et al. [5] requires cropped images and tedious data augmentation. Landmarks can also be detected by segmenting the relevant vertebrae [2]. Our proposed model is totally end-to-end, requiring no pre-processing, and is fully automatic, eschewing any hand-crafted feature extractions.

2 Method

Given an X-ray image, we formulate the landmark detection problem as identifying n landmarks localizing the relevant vertebrae. Each training image x_i, for $i = 1, \ldots, m$, is annotated by an associated $2n$-dimensional landmark vector y_i. Through supervised learning, a CNN can be trained to extract landmarks automatically, by minimizing the standard mean squared error (MSE) loss

$$\text{MSE} = \frac{1}{m} \sum_{i=1}^{m} (y_i - \hat{y}_i)^2, \tag{1}$$
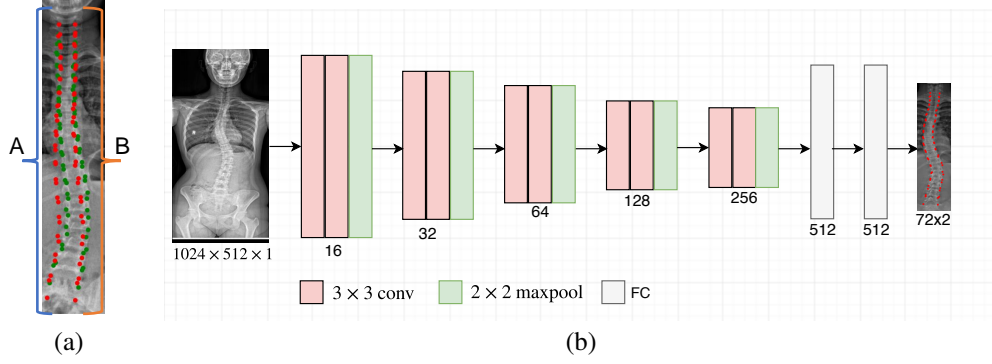
Figure 1: Convolutional neural networks for landmark detection from spine X-ray images: (a) Illustration of the bipartite distance in a spinal image based on the ground truth (green) and predicted (red) landmarks. (b) Model architecture.

where y_i are the ground-truth landmarks and $\hat{y}_i$ are the predicted landmarks. However, the MSE loss ignores inter-landmark relationships. To guide a CNN in the detection of landmark coordinates while learning spinal shape, we propose a novel distance measure—bipartite distance.

Referring to Figure 1a, we regard the ground-truth (green) landmarks A_y on the left and B_y on the right of the spine as the two disjoint sets of vertices of a complete bipartite graph whose edges connect every landmark in A_y with all landmarks in B_y. The same holds for the predicted (red) landmarks, $A_{\hat{y}}$ and $B_{\hat{y}}$. This leads to a shape-aware loss, which penalizes the CNN model when the pairwise distances between the predicted landmarks deviate from those between the ground truth landmarks. Letting d^e denote the Euclidean distance between ground-truth landmarks connected by edge e of the graph and $\hat{d}^e$ denote the Euclidean distance between the corresponding predicted landmarks, the bipartite distance (BPD) is

$$\text{BPD} = \frac{1}{m} \sum_{i=1}^{m} \sum_{e} \left| d_i^e - \hat{d}_i^e \right|. \tag{2}$$

We employ the loss function

$$L = \text{MSE} + \alpha \, \text{BPD}, \tag{3}$$

where α weighs the BPD term against the MSE.

3 Implementation Details

Our dataset consists of 100 high-resolution anterior-posterior spinal X-ray images with signs of mild to severe scoliosis. Since the cervical vertebrae are seldom involved in spinal deformity and the identification of the bottom cervical vertebra could be important, we selected 18 vertebrae: C7 (cervical), T1–T12 (thoracic), and L1–L5 (lumbar). Medical experts provided binary segmentation annotation by labeling the $n/4 = 18$ vertebrae in the X-ray images. The 4 corners of each vertebral region serve as landmarks. They were automatically extracted by applying FAST [3] to the expert-segmented labels. Therefore, associated with each spinal image are 72 landmarks to be estimated.

As shown in Figure 1b, our model is a CNN comprising five convolutional layers and three fully-connected (FC) layers. Leaky-ReLU is used as the activation function in each layer. The convolutional layers have feature sizes 16, 32, 64, 128, and 256. In each layer, two 3×3 convolution operations are followed by a 2×2 maxpooling layer. After every convolutional layer, we use a batch-normalization layer and a dropout layer with the rate of 0.25. After two FC layers with 512 neurons, a final FC layer of $2n = 144$ neurons is used to produce the image-plane coordinates of the landmarks. The model is implemented in Tensorflow with Python 3 and runs on a Tesla P40 GPU on a 64-bit Intel(R) Xeon(R) 440G CPU.

The dataset was split into training (80 images), testing (15 images), and validation (5 images) sets. All the images were resized to $1024 \times 512 \times 1$ and normalized to $[0, 1]$ before feeding them to the network. When the model is trained using our MSE-BPD loss, we used $\alpha = 0.01$ in (3). As a
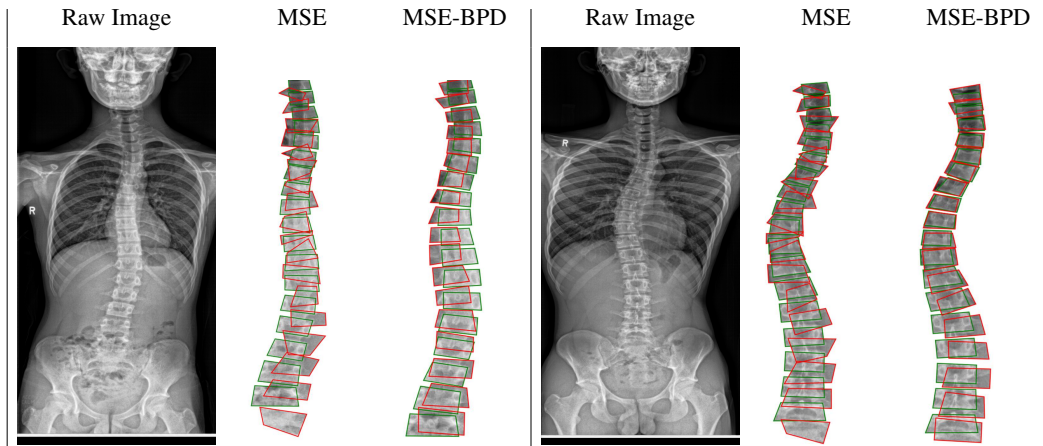
Figure 2: Qualitative comparison shows improved spinal shape and landmark detection performance by our model (MSE-BPD loss) relative the baseline model (MSE loss) in two spinal X-ray images from the test set. Green boxes bound vertebrae based on the ground-truth landmarks; red boxes bound vertebra based on the model-predicted landmarks.

baseline, we trained the same architecture using only the MSE loss; i.e., $\alpha = 0$. The models were trained with a minibatch size of 4. We used the Adam optimizer with a learning rate of 0.0001 and momentum 0.9.

4   Experimental Results

We compared the performance of the proposed model (MSE-BPD) against the baseline (MSE) both qualitatively and quantitatively. Qualitative comparisons (e.g., Figure 2), show better agreement of our model over the baseline model. The irregular spinal shapes in the baseline model are mitigated by our model. Moreover, the landmark detection performance is also improved in our model, which achieves a correlation score of 0.95 compared to the baseline model's score of 0.92 (Pearson correlation coefficient). Moreover, one-way ANOVA analysis confirms that the landmarks predicted by our model have no significant difference with the ground truth landmarks (*p-value* < 0.05).

5   Conclusions

The detection of vertebral landmarks is crucial for the accurate measurement of scoliosis in spinal X-ray images. To this end, we proposed a new loss function which guides the training of a CNN vertebral (corner) landmark detection model to perform reliable shape-aware predictions.

References

[1] Ebrahimi, S., Gajny, L., Skalli, W. and Angelini, E. [2019], 'Vertebral corners detection on sagittal x-rays based on shape modelling, random forest classifiers and dedicated visual features', *CMBBE: Imaging & Visualization* **7**(2), 132–144.

[2] Imran, A.-A.-Z., Huang, C., Tang, H., Fan, W., Cheung, K. M., To, M., Qian, Z. and Terzopoulos, D. [2019], End-to-end fully automatic segmentation of scoliotic vertebrae in spinal X-ray images, *in* 'Medical Imaging Meets NeurIPS Workshop', Vancouver, Canada.

[3] Rosten, E. and Drummond, T. [2006], Machine learning for high-speed corner detection, *in* 'European Conference on Computer Vision', Springer, pp. 430–443.

[4] Weinstein, S. L., Dolan, L. A., Cheng, J. C., Danielsson, A. and Morcuende, J. A. [2008], 'Adolescent idiopathic scoliosis', *The Lancet* **371**(9623), 1527–1537.

[5] Wu, H., Bailey, C., Rasoulinejad, P. and Li, S. [2017], Automatic landmark estimation for adolescent idiopathic scoliosis assessment using BoostNet, *in* 'Proc. 2017 MICCAI Conf.'.