

A Implementation Details and Ablation

Regression Network of Spine Curves. The spine presents elongated and thin curve structures which makes mathematical modelling a challenging task. Attention-based ‘Transformer’ models have become increasingly used in natural language processing tasks, thanks to their ability to learn effective feature representations. In this work we adopt the BoTNets architecture. The standard spatial 3×3 convolutional layer is replaced by Multi-Head Self-Attention (MHSA). It also has 1.2x fewer parameters compared to ResNet50 (20.8×10^6 vs 25.5×10^6). In line with our work, ¹ showed that a lightweight transformer model using significantly less parameters also surpasses CNN-based and Vision Transformer (ViT) models on the Synapse dataset, SegPC, and ISIC 2017 dataset.

Ablation Experiments. We experiment with varying input sample size during training with an overall improved test performance for spine mask prediction of +3.4 IoU using the whole training set of 30k available versus 500 samples. Therefore, the network benefits from training on large datasets and it improves its ability to generalise. We also experiment adding more than one transformer layer on top of ResNet50. This does not significantly boost performance on the order of +0.002px and +0.003px average improvement for coronal and sagittal curve regression respectively.

B 3D Spine Curve from 2D

3D Spine from 2D Curves. In this section, we outline the steps taken for 3D spine shape recovery from two 2D planes i.e. coronal and sagittal. Traditionally the planes used for estimating the projections of 2D planes to 3D with multi-view stereo images have small rotation angles. In the case of orthogonal planes (coronal and sagittal), MVS does not work and instead the reconstruction is computed. The output of our network are 2D spine curves on the coronal (XY) plane and sagittal (YZ) plane. The 3D spine curve is reconstructed directly from the curves in the two orthogonal planes (see Figure 1).

3D Spine Shape Rotation 360° . We show the 360° rotation around the z-axis of predicted 3D spines for a normal, and severe scoliosis case. To obtain the 3D spine masks from two 2D segmentation, we use the 4 points in antero-posterior and left-right to generate a series of bounding ellipse as we go down in z. The spines in the video are obtained by rotating along the z-axis. Our model works well in estimating sagittal MRI projections for normal spines and even severe scoliosis spines.

¹ Heidari, M., Kazerouni, A., Kadarvish, M.S., Azad, R., Aghdam, E.K., Cohen-Adad, J., Merhof, D.: Hiformer: Hierarchical multi-scale representations using transformers for medical image segmentation. WACV pp. 6191– 6201 (2022)

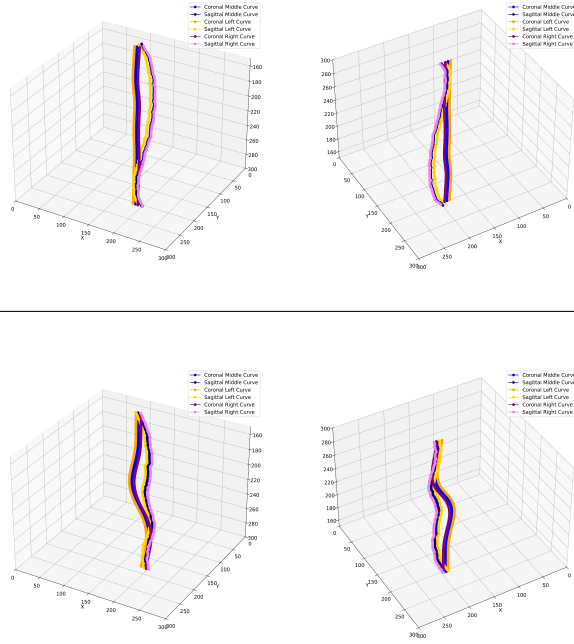


Fig.1. 3D Spine From 2D. We show on the same plot *top left* the coronal (XY) and sagittal (YZ) curves of the same patient. We show the same plot in *top middle* for 90° rotation and elevation angle 30. We show the 3D spine curve rotating around the z-axis *top right* obtained from the 2D curves. *Bottom* is same as *top* row but for a severe scoliosis case with important lateral deviation on the [XY] plane.