Supplementary 1: Development and Training of the Model.

Two separate neural networks were developed in our system to perform rib segmentation and labelling, and rib fracture detection, respectively. The obtained rib segmentation mask was used to confine the search range for rib fractures, and rib labels were to identify on which rib the detected fracture was located.

1. Rib segmentation and labelling

We used a fully convolutional neural network (FCNN) based approach as described in the reference [1] for rib segmentation and labelling. But instead of only classifying the rib centerline voxels in the CT scans, we targeted separating out the voxels of the whole rib in this work. Regarding the FCNN network architecture, we adopted the Foveal network [2] which comprises of three different resolution layers that combine both low-level features to enable precise localization and high-level features to capture semantic context. First, the network was applied to generate a multi-label probability map for the first pair of ribs, the twelfth pair of ribs, the collection of all intermediate ribs, and the background. Then the ribs were segmented and labelled from the probability map by iterative tracing and counting algorithm as described in the reference [1].

2. Rib fracture detection

Inspired by the success of general object detection from natural 3-colour-channel images[3], we employed the cascade Faster R-CNN network to detect rib fractures from 3D CT scan images. Unlike the general objects in RGB images, the appearance of rib fractures does not exhibit scale invariance and therefore we dropped the ROI pooling layer and only kept the 3D region proposal network, as shown in Figure S1. The network took the CT as input, and output a set of bounding boxes each encompassing the detected rib fracture with a confidence score. Those bounding boxes were processed with the Non- Maximum Suppression (NMS) algorithm to get rid of spatially closely clustered boxes.

Similar to the Faster R-CNN[3], we also used the anchor boxes for the training of our detection network that served as references at multiple scales and aspect ratios of rib fractures. In this work, we used 3 scales with width of 3 mm, 6 mm and 12 mm, and 7 different aspect ratios of 1:1:1, 2:1:1, 1:2:1, 2:2:1, 1:1:2, 2:1:2 and 1:2:2. An anchor box was assigned a positive label if it has the Intersection- over-Union (IoU) higher than 0.7 with any ground-truth box which was manually labelled by the radiologists around the rib fractures they found or has the highest IoU with a ground-truth box; whereas we assigned negative labels to a non-positive anchor box if it has IoU lower than 0.3 with any ground-truth box. Therefore, multiple anchor boxes are allowed to be positive corresponding to a single ground-truth box and those anchor boxes that are neither negative nor positive won't be used in training.

We adopted the same weighted multi-task loss function as defined in the reference [3]:

$$L(\{p_i\},\{t_i\}) = \frac{1}{N_{cls}}\sum_i L_{cls}(p_i,p_i^*) + \lambda\frac{1}{N_{reg}}\sum_i p_i^* L_{reg}(t_i,t_i^*).$$

But different from the reference [1], the classification loss $\boldsymbol{L_{cls}}$ is defined as the focal loss in the reference [4] over two classes (rib fracture vs. not rib fracture) to address the problem where the samples of two classes are highly unbalanced. The same smooth L1 loss function defined in the reference [3] is still used as the regression loss. The network was trained with a variant of the standard stochastic gradient descent algorithm named Adam [5].

3. Dataset

We collected 3,382 CT thoracic scan data from 10 different hospitals, of which 2,855 scans have at least one rib fracture annotated, with a total of 15,082 rib fractures annotated from all the data. Each ground truth bounding box was labelled by one or more senior radiologists using a specific annotation software tool. The dataset was randomly split into a training set of 3,125 scans and a validation set of 257 cases for the training of the rib fracture detection network. Out of the 3,125 training data, a subset of 800 scans were randomly selected and manually annotated for training the rib segmentation and labelling network. The age and gender distribution of the patients of these collected 3,382 CT data were listed in Table S1.

4. Reference

[1]Matthias Lenga, Tobias Klinder, Christian Bürger, Jens von Berg, and Astrid Franz, Cristian Lorenz. Deep Learning Based Rib Centerline Extraction and Labeling. arXiv:1809.07082, 2018

[2]T. Brosch and A. Saalbach, "Foveal fully convolutional nets for multi-organ segmentation",Proc. SPIE Medical Imaging 2018: Image Processing, vol. 10574, 105740U;doi: 10.1117/12.2293528; https://doi.org/10.1117/12.2293528

[3]Ren S, Girshick R, et al. Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks[J]. IEEE Transactions on Pattern Analysis & Machine Intelligence, 2017, 39(6):1137-1149.

[4]Lin T Y, Goyal P, Girshick R, et al. Focal Loss for Dense Object Detection[J]. IEEE Transactions on Pattern Analysis & Machine Intelligence, 2017, PP(99):2999-3007.

[5]Kingma, D. P., & Ba, J. L. Adam: a Method for Stochastic Optimization. International Conference on Learning Representations, 2015, 1–13.

Table S1. Age and gender distribution of the patients of the collected dataset.

Sex	Age							
	20~30	30~40	40~50	50~60	60~70	70~80	80~90	90~100
male	92	241	497	610	540	221	87	14
female	20	77	149	300	262	163	96	13
total	112	318	646	910	802	384	183	27

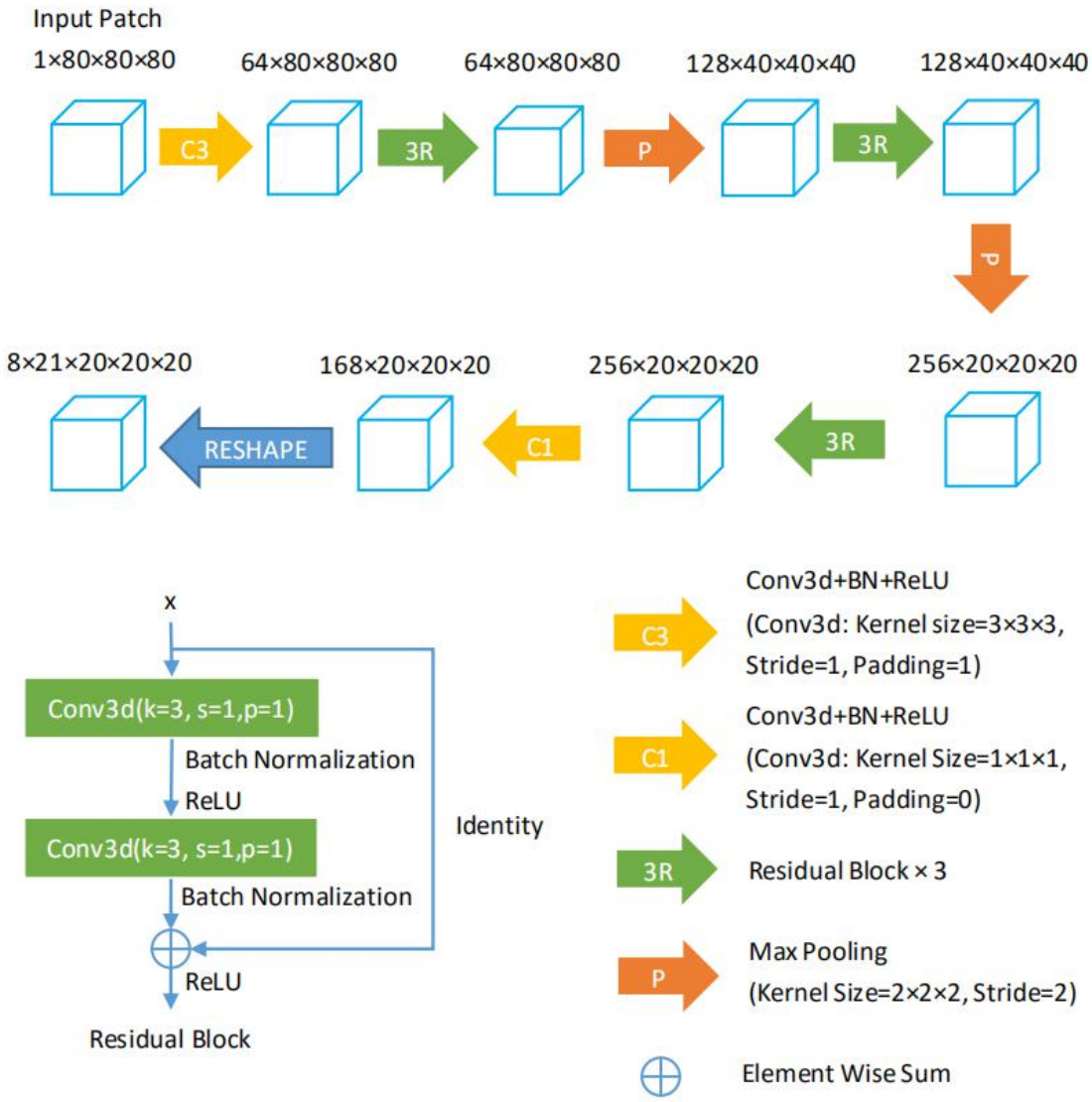

Figure S1. 3D Region Proposal Network architecture. An input patch size of 80×80×80 was adopted in our experiment, which was shown in the sketch as a box. In the training stage, fix sized image patches are fed into the network and it outputs a set of cuboid object proposals for each patch, which are used to calculate the weighted combined loss with anchors generated from labelled rib fracture boxes. BN: batch normalization; Conv3d: convolution network 3D; ReLU: rectified linear unit.

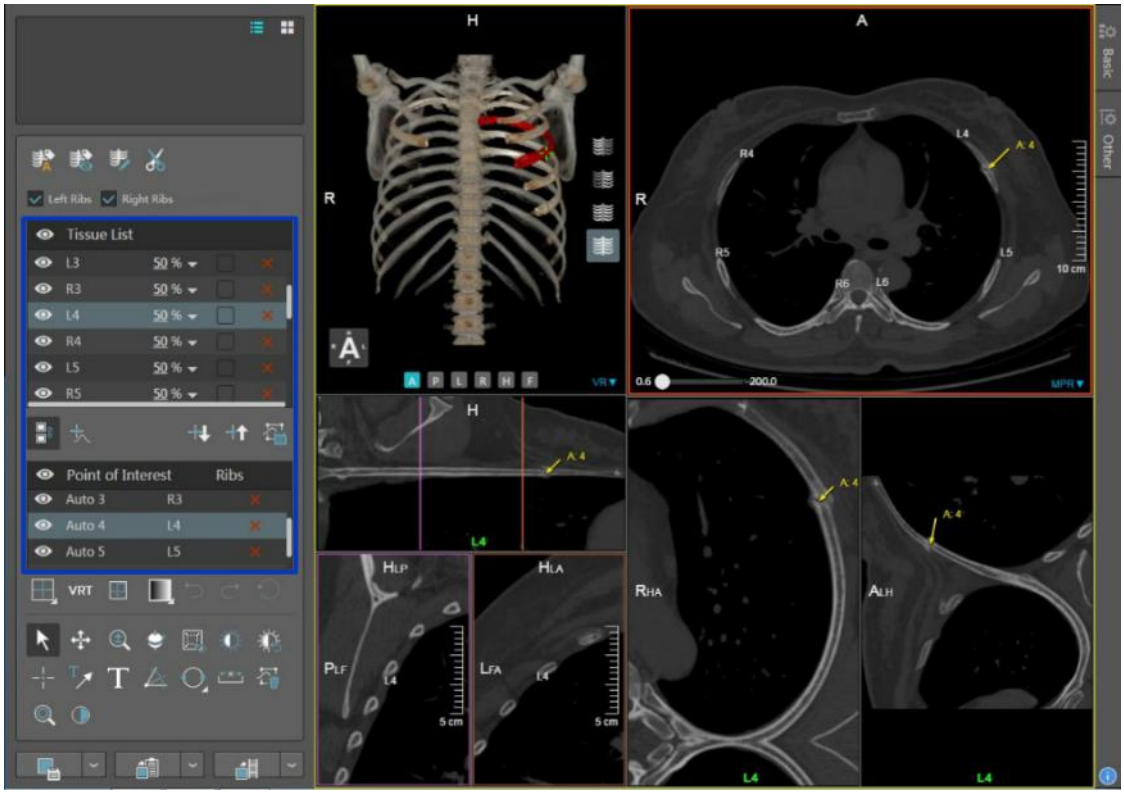

Figure S2. Representative working view of the prototype DL-based rib fracture detection tool. The rib fractures detected by DL were listed in a navigation window on the left panel highlighted in blue. The findings were also marked by the yellow arrows next to the fractures in the axial view. The VR, straight CPR, and CPR views were automatically generated for the readers to check the existence of the fracture.