Appendix 1

Data sets

Out of all 1734 cases, 213 cases with a history of spinal surgery, 26 with severe spinal deformities, eight with metabolic bone disease, 15 cases with spinal tuberculosis, 12 cases with spinal tumors, 24 cases with poor image quality, and 68 cases with severe osteophytes were excluded based on the study's study's inclusion and exclusion criteria. Therefore, 1,368 anteroposterior digital radiographs of the lumbar spine were included, which were randomly assigned as training set, validation set, and test set according to the ratio of 65%, 20%, and 15%, respectively. The training set was used to optimize the model parameters (n=893), the validation set was used to adjust the model hyperparameters (n=269), and the test set was used to verify the performance of the trained network (n=206). The training and test sets did not overlap to ensure an objective evaluation of the model.

The standard and equipment of anteroposterior digital radiography of the lumbar spine

Gansu Provincial Hospital of Traditional Chinese Medicine imaging center equipment: U.S. RICO (Kodak DR3000), GE (GE Definium6000).

Photographic criteria: The patient stood in an anteroposterior position in front of the photographic frame with the centerline aligned approximately 5 cm above the anterior superior iliac spine (at the level of the L3 vertebrae). The radiation field was 28×35.0 cm, the photographic distance was 120 cm, the exposure was automatically controlled by the central ionization chamber, the filter-grid ratio was 10:1, the voltage was 60–70 kV, and the current was 20–40 mAs. The patient was exposed by holding his/her breath in a calm breathing state.

Appendix 2

Model construction

The Cascaded DARK model construction process was divided into two stages to retain detailed information on lumbar spine images: lumbar spine vertebrae detection and landmark identification. In the first stage, the HRNet output 20 channels, and each channel corresponds to a key point responsible for detecting the coarse position of each vertebra. According to the vertebrae to which each landmark belongs, the outer rectangle expanded by 60 pixels of every four landmarks determines the region where a vertebra is located. In the second stage, the output layer of HRNet has four channels, which are used to more precisely identify the four key points on the vertebrae from L1 to L5. The above two phases consist of two DARK models, denoted as Cascaded DARK for convenience, both of which use the same network architecture HRNet-32 and differ only in the number of output channels. The same training strategy described in DARK is used to train both models independently. HRNet's HRNet's main feature is that the feature maps maintain their high resolution throughout the process by gradually adding multiple branches of low-resolution feature maps in parallel to the main network of high-resolution feature maps, allowing for multi-scale fusion and feature extraction. The final estimated key points are the output of the high-resolution backbone network; therefore, the method can maintain high resolution from beginning to end and achieves strong semantic information and accurate localization for key points.

Heatmap technology is widely used in landmark detection tasks (37). We used the heatmap method to supervise the recognition of neural networks to the key points of the vertebral body. Furthermore, heatmap, similar to class label smoothing regularization, may effectively reduce the risk of model overfitting in training by taking into account not only contextual clues but also the inherent target position ambiguity. The coordinate encoding was a label representation process from the coordinate to the heatmap, which generated a 2-dimensional Gaussian distribution centered at the labeled coordinate of each landmark. Formally, we denote by $\boldsymbol{g}=(u,v)$ the ground-truth coordinate of a landmark. The resolution reduction was defined as:

$$\boldsymbol{g}' = (u', v') = \frac{\boldsymbol{g}}{\lambda} = \left(\frac{u}{\lambda}, \frac{v}{\lambda}\right) \quad [1]$$

where λ is the downsampling ratio.

© Quantitative Imaging in Medicine and Surgery. All rights reserved.
https://dx.doi.org/10.21037/qims-22-540

Subsequently, the heatmap centered at the accurate coordinate $\boldsymbol{g}'$ can be denoted as:

$$\varsigma(x, y; \boldsymbol{g}') = \frac{1}{2\pi\sigma^2} \exp\left(-\frac{(x-u')^2 + (y-v')^2}{2\sigma^2}\right) \quad [2]$$

where (x, y) specifies a pixel location in the heatmap, and σ denotes a fixed spatial variance. In this work, σ is set to 2.

Coordinate decoding is a process from heatmap to coordinate, inferring the underlying maximum activation based on the distribution structure of the predicted heatmap. At test time, the HRNet outputted multi-channel heatmaps where each heatmap corresponded to the probability of a landmark. The DARK method and an efficient Taylor-expansion-based coordinate decoding procedure take the HRNet-predicted heatmaps as input and output precise landmark coordinates in the original image space.

Finally, the mean square error was selected as the loss function, and the Adam optimizer was adopted. The parameters obtained by pre-training on the ImageNet dataset were used for initializing the weights of HRNet through transfer learning techniques to accelerate the convergence of the model and reduce the risk of overfitting. During the prediction process, the exact 20 landmark locations can be obtained from the original radiographs with the trained Cascaded DarkPose model following the key point localization process. Based on the calculation method in parameter measurement, the relevant radiological parameters are then calculated from the coordinates of the predicted landmarks, thus enabling automatic measurement.

References

37. Leonardi R, Giordano D, Maiorana F, Spampinato C. Automatic cephalometric analysis. Angle Orthod 2008;78:145-51.

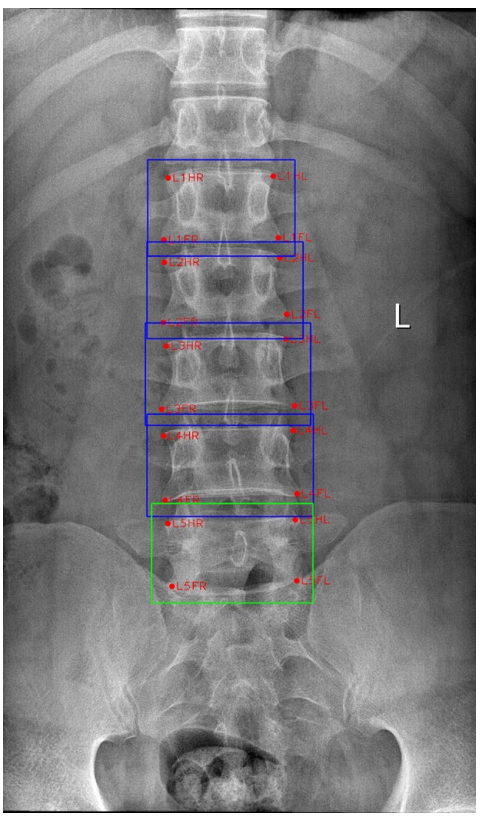

Figure S1 The bounding box of vertebral bodies.

© Quantitative Imaging in Medicine and Surgery. All rights reserved.
https://dx.doi.org/10.21037/qims-22-540

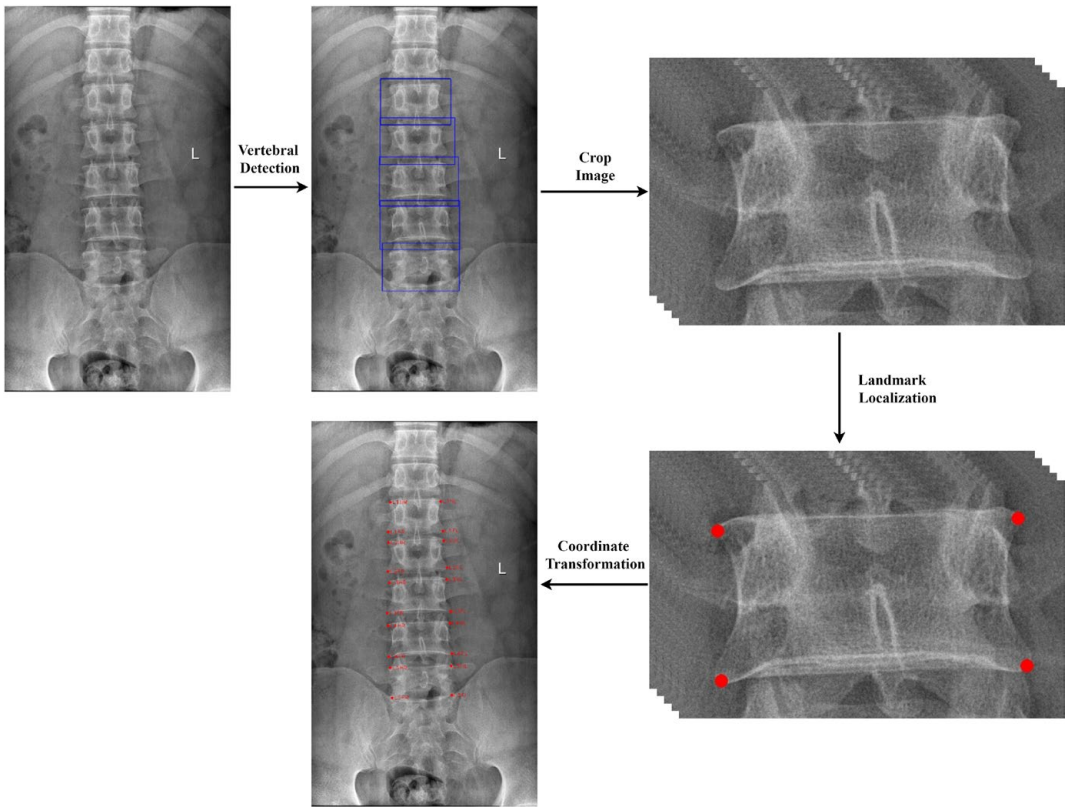

Figure S2 Key point positioning process.

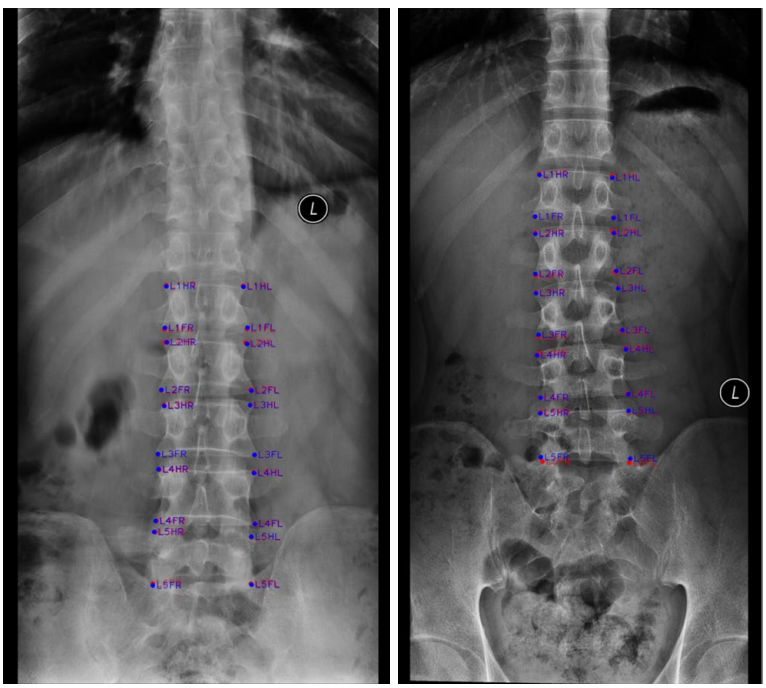

Figure S3 Representative images illustrate landmark detection produced by our model. The blue points are the reference standard, and the red points are the model prediction point.

© Quantitative Imaging in Medicine and Surgery. All rights reserved.
https://dx.doi.org/10.21037/qims-22-540