Received: 4 May 2022 | Revised: 27 July 2022 | Accepted: 24 August 2022

CAAI Transactions on Intelligence Technology

DOI: 10.1049/cit2.12137

ORIGINAL RESEARCH

Lumbar spine localisation method based on feature fusion

Yonghong Zhang[1,2] | Ning Hu[3] | Zhuofu Li[4,5,6] | Xuquan Ji[2,7] | Shanshan Liu[4,5,6] | Youyang Sha[8] | Xiongkang Song[1,2] | Jian Zhang[1,2] | Lei Hu[1,2] | Weishi Li[4,5,6]

[1]Robotics Institute, School of Mechanical Engineering and Automation, Beihang University, Beijing, China

[2]Beijing Zhuzheng Robot Co., LTD, Beijing, China

[3]Department of Mechanical, Aerospace and Biomedical Engineering, University of Tennessee, Knoxville, Tennessee, USA

[4]Department of Orthopaedics, Peking University Third Hospital, Beijing, China

[5]Engineering Research Center of Bone and Joint Precision Medicine, Ministry of Education, Beijing, China

[6]Beijing Key Laboratory of Spinal Disease Research, Beijing, China

[7]School of Biological Science and Medical Engineering, Beihang University, Beijing, China

[8]Department of Computer Science, University of Warwick, Coventry, UK

Correspondence

Weishi Li, Department of Orthopaedics, Peking University Third Hospital, No. 49 North Garden Road, Haidian District, Beijing 100191, China.
Email: puh3liweishi@163.com

Funding information

Beijing Natural Science Funds-Haidian Original Innovation Joint Fund, Grant/Award Number: L202010; National Key Research and Development Program of China, Grant/Award Number: 2018YFB1307604

Abstract

To eliminate unnecessary background information, such as soft tissues in original CT images and the adverse impact of the similarity of adjacent spines on lumbar image segmentation and surgical path planning, a two-stage approach for localising lumbar segments is proposed. First, based on the multi-scale feature fusion technology, a non-linear regression method is used to achieve accurate localisation of the overall spatial region of the lumbar spine, effectively eliminating useless background information, such as soft tissues. In the second stage, we directly realised the precise positioning of each segment in the lumbar spine space region based on the non-linear regression method, thus effectively eliminating the interference caused by the adjacent spine. The 3D Intersection over Union (3D_IOU) is used as the main evaluation indicator for the positioning accuracy. On an open dataset, 3D_IOU values of 0.8339 ± 0.0990 and 0.8559 ± 0.0332 in the first and second stages, respectively is achieved. In addition, the average time required for the proposed method in the two stages is 0.3274 and 0.2105 s respectively. Therefore, the proposed method performs very well in terms of both precision and speed and can effectively improve the accuracy of lumbar image segmentation and the effect of surgical path planning.

KEYWORDS

CT image, lumbar spatial orientation, multi-scale information fusion

1 | INTRODUCTION

Lumbar spine disease is a very common type of spinal disease that is often associated with pain. In particular, lumbar spinal stenosis (LSS) is a very common disease of the lumbar spine that can lead to back and lower limb pain, mobility problems and other disabilities [1]. In the United States, LSS is the most common reason for spinal surgery in people over the age of

Yonghong Zhang, Ning Hu and Zhuofu Li contributed equally to this work.

65 [2]. Computed tomography (CT) images have irreplaceable advantages in the diagnosis and treatment of LSS, particularly with the rapid development of artificial intelligence and robot technology, the autonomous planning of robotic surgical path based on preoperative CT images has become a research hotspot [3, 4]. The autonomous planning of the robotic surgical path requires accurate localisation, segmentation and 3D reconstruction of the lumbar. However, due to the large number of CT image data and the interference of the soft tissue, thoracic vertebra, pelvis and other useless background information, it is very difficult to segment the lumbar spine directly from the original image. Therefore, a feasible and effective method to accurately locate the lumbar region before image segmentation of the lumbar spine is required.

Several studies have demonstrated the effectiveness of this method. Janssens et al. developed cascaded 3D Fully Convolutional Networks (FCNs), which first locates the whole region of the lumbar spine to achieve accurate segmentation of the lumbar spine [5]. Sekuboyina et al. employed a multi-layered perceptron performing non-linear regression to locate the lumbar region using the Global context and achieved precise segmentation of the five lumbar vertebrae [6]. Both methods locate the whole region of the lumbar spine, and directly segment each segment of the spine as different targets for image segmentation. Although, this method can effectively reduce the interference from unnecessary background information such as soft tissue, it still has the problem of poor effect because of the large amount of information in the whole lumbar region and the high degree of similarity between the spinal segments. As shown in Figure 1, L3 and L4 are prone to ambiguity and segmentation errors due to their high structural similarity.

In order to solve the above problems, Payer et al. developed a three-stage automatic segmentation method of the spine, namely the first positioning for the whole area of the spine, then using the heatmap regression algorithm to detect and position the landmarks of each spine. Finally, a series of large enough bounding boxes with a fixed size were used to surround each section of the spine, and the precise segmentation of each section of the spine is realised in this bounding box [7]. However, this method did not consider compression fractures and the differences in spinal morphology of each patient, resulting in inadequate robustness. Lessmann et al. determined the spatial position of each spine and effectively eliminated the interference of adjacent vertebrae by assigning corresponding logical rules to the algorithm, sliding sampling in the CT image using a sliding window, and determine whether there is a complete spinal segment within the window [8]. Although the method is valid, the complicated logical operation process and sliding sampling results in low efficiency.

The spatial localisation of lumbar vertebrae is similar to target detection in natural image processing. In recent years, target detection has become a research hotpot. In the field of two-dimensional natural image processing, target detection

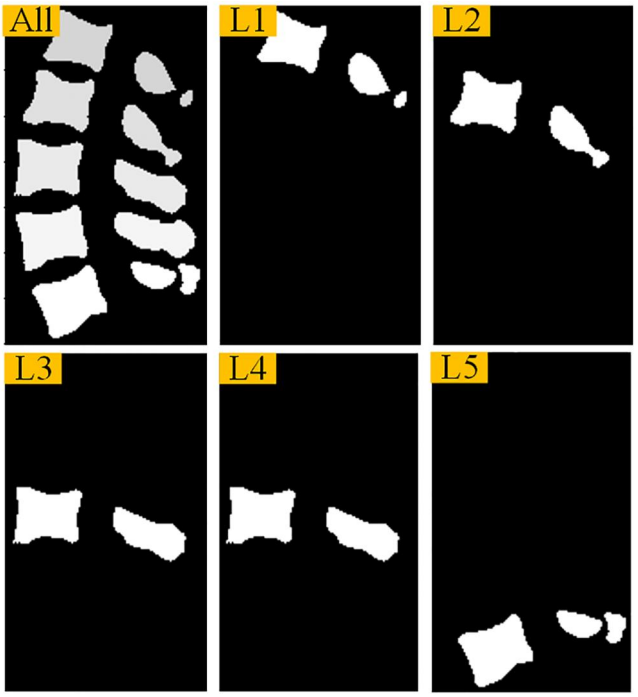

FIGURE 1 Effect of structural similarity on lumbar image segmentation

algorithms, such as RCNN series [9–11] and Yolo series [12–15] have emerged. However, there are few studies in the field of medical image processing, particularly in the field of three-dimension medical image processing. At present, most of the related studies are based on the improvement of the natural image object detection algorithm. Vos et al. proposed a method of 2D slice target localisation based on multiple directions to achieve 3D target localisation [16]. However, this method loses 3D spatial information and requires multiple positioning, making less accurate and efficient. Xu et al. proposed a CT image organ localisation algorithm based on Faster RCNN [11], the method is completely based on 3D operation, which is more consistent with the high-dimensional features of CT images [17]. However, this method produces a large number of anchor boxes during operation, and the final target area needs to be obtained through post-processing, resulting in low operation efficiency.

To efficiently and accurately locate the spatial region of each lumbar vertebra, we propose a two-stage method; in this approach, the overall spatial region of the lumbar vertebra is located first, and the spatial region of each lumbar vertebra within the overall region is subsequently determined. This method fully combines the particularity of lumbar number fixation in lumbar positioning task. Based on multi-scale feature fusion technology, it adopts the method of non-linear regression output coordinate information of the target area directly; it also does not produce candidate boxes, and there is no need to obtain the target area via post-processing. Hence,

this approach offers a significant advantage in terms of the positioning precision and efficiency. In addition, we propose a new loss function, which can greatly improve the training effect of the algorithm.

The remainder of this paper is organised as follows. Section 2 introduces the two-stage method and the loss function in detail. Section 3 details the network training. Section 4 presents the evaluation of the validity and feasibility of the proposed method through comparative algorithm experiments. The fifth part will discuss our research. Finally, we will summarise our work.

2 | METHODS

As illustrated in Figure 2, we used two stages to achieve precise localisation of the lumbar spatial region. In the first stage, the lumbar rough positioning network (LRP-Net) used multi-scale feature fusion and non-linear regression to process raw CT images to achieve precise localisation of the global region of the lumbar spine, so as to effectively eliminate useless background information such as soft tissue in the original CT image. In the second stage, the lumbar multitarget detection network (LMD-Net) was used to accurately locate each lumbar spatial region within the global region of the lumbar spine using non-linear regression methods. Through the above method, each lumbar vertebra can be accurately intercepted from the original CT image, providing a good foundation for image segmentation and independent surgical path planning.

2.1 | Overall regional positioning of lumbar spine

We used LRP-Net to locate the entire lumbar spine region in the original CT image. For the convenience of expression, the whole lumbar region is represented by ROI_{Lumbar} (a cuboid). This cuboid size is determined by the body's diagonal coordinates and can be expressed as (z_min, z_max, y_min, y_max, x_min and x_max). Thereafter, we performed direct regression on ROI_{Lumbar} to obtain six coordinate data. The LRP-Net structure is illustrated in Figure 3.

LRP-Net consists of two sections: the Backbone network and the 3D-Roi Regression network. The Backbone network is mainly used to extract the features of CT images, which is improved on U-Net [19] under the inspiration of Payer et al. [7]. The 3D-Roi Regression network conducts a series of processing on the feature map output by the Backbone network, and finally outputs ROI_{Lumbar} through non-linear regression.

The encoder of the Backbone network consists of four layers, each layer includes two CBR (Convolution + BatchNorm + ReLU) modules and one CBRP

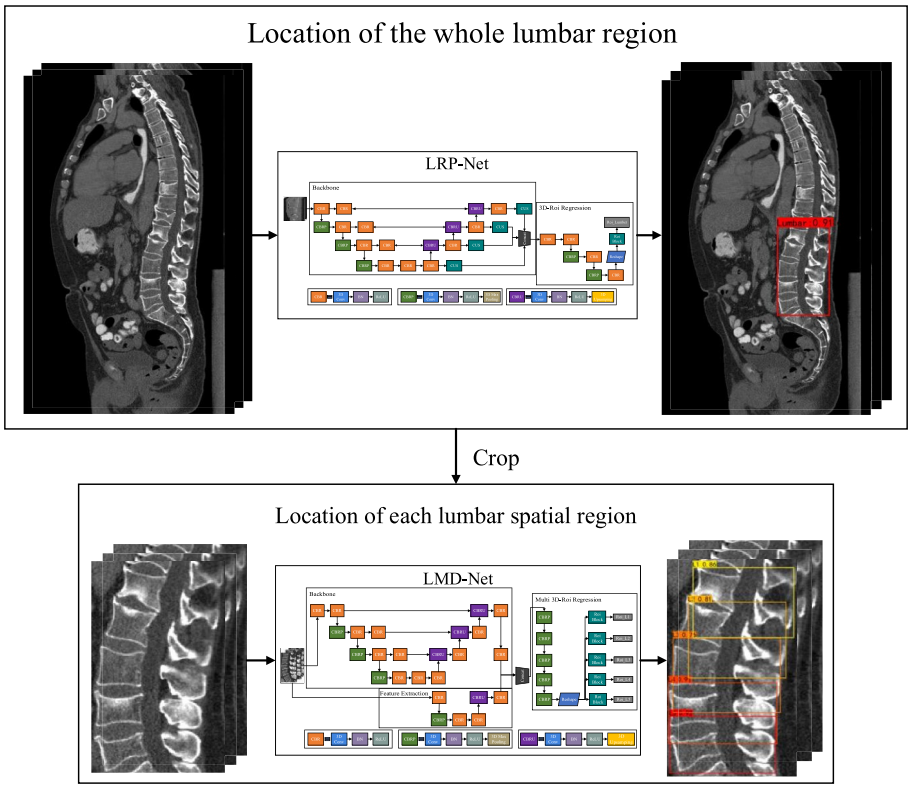

FIGURE 2 Overall plan of lumbar spine positioning

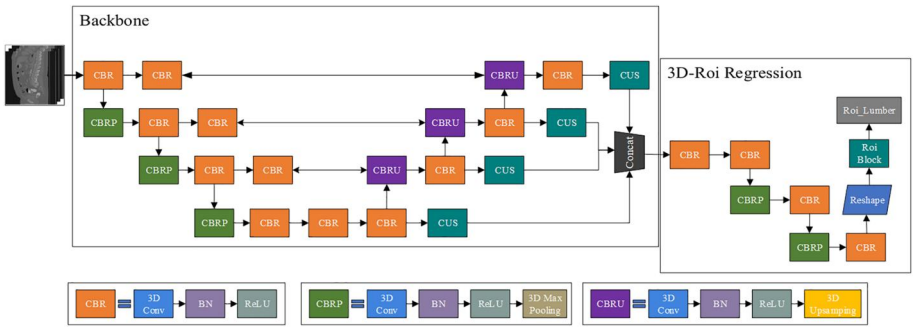

FIGURE 3 Structure diagram of LRP-Net

FIGURE 4 Our Continuous Upsampling (CUS) module automatically obtains the required continuous upsampling times according to the input and output sizes, and ensure that the input feature images reach the target size through continuous upsampling. The CUS module contains N − 1 upsampling modules with an upsampling multiplier of two and an adaptive fine-tuning upsampling module

(Convolution + BatchNorm + Re-LU + Pooling) module. To extract features more effectively, we added the CBR module before Max Pooling and formed the CBRP module.

In the decoder, each layer includes a CBRU (Convolution + BatchNorm + ReLU + Upsampling) module, a CBR module and a CUS(Continuous Upsampling) module. Similarly, to extract features more effectively, we add a CBR module for feature extraction before the upsampling operation. The encoder extracts shallow features, which contain sufficient spatial location information; by contrast, the decoder extracts deeper semantic information. Through feature channel stacking, deep and shallow feature information can be fused, and a fusion feature map with rich spatial location and semantic information can be obtained. To prevent the channel dimension from increasing, we reduce the number of feature channels and feature fusion through a three-dimensional convolution operation after splicing. Because the feature map size of each layer in the decoder is different, feature fusion cannot be realised directly via stacking on the dimension of feature channel. To preserve the information in each feature map, we first need to increase the size of the smaller feature maps and then perform the stack operation once all images are unified. We added the CUS module in Backbone network, which can quickly realise the size amplification of each layer's feature map of the decoder through continuous upsampling and also unify the size of each layer to the same size as that of the input image. The corresponding structure is presented in Figure 4.

To prevent feature information loss caused by continuous up-sampling, a three-dimensional convolution operation is performed on the amplified feature graph obtained from up-sampling, and the number of output feature channels is set to CUS_Count. In the 3D-Roi Regression network, we arrange the CBR module and CBRP module alternately, to obtain concise feature information and realise the shrinkage of feature images; notably, this reduces the computational complexity in the subsequent non-linear regression of all connected layers. Finally, we realise the non-linear regression through three fully connected layers and determine ROI_{Lumbar}.

2.2 | Spatial orientation of each lumbar vertebrae

We used ROI_{Lumbar} to crop the original CT image to obtain a local image that includes the whole lumbar spine and serves as input to LMD-Net. We represent these five lumbar regions as follows: ROI_{L1}, ROI_{L2}…ROI_{L5}. These five areas can also be described in the same manner as ROI_{Lumbar}. In the second stage, we treat the spatial positioning of five lumbar vertebrae as a regression problem, and LMD-Net obtains the respective coordinates of ROI_{L1}, ROI_{L2}…ROI_{L5} by non-linear regression. The LMD-Net structure is illustrated in Figure 5.

The LMD-Net network includes three parts: the Backbone, Feature Extraction and Multi 3D-Roi Regression networks. Among them, the Backbone network can effectively extract the deep features of local CT images, to realise the general spatial orientation of the five lumbar spines. The Feature Extraction network is responsible for extracting the shallow features of local CT images. Thereafter, pixel superposition and fusion are carried out on the feature map output by the Backbone and Feature Extraction networks. The Multi 3D-Roi Regression network processes the fusion feature images, so as to obtain ROI_{L1}, ROI_{L2}…ROI_{L5}.

LMD-Net's Backbone network is roughly the same as that of LRP-Net, albeit with two main differences:

(1) The fusion of the feature images corresponding to the encoder and decoder is realised via pixel-by-pixel addition, rather than channel stacking. Owing to the higher complexity of LMD-Net and the small difference in the sizes of the input images for the two networks, the hardware level will be higher if the feature channel dimensions are stacked. Hence, we chose to add them in a pixel-by-pixel manner to achieve feature fusion.
(2) The size and pixel superposition of the four-layer feature maps during feature fusion are not uniform. In LMD-Net,

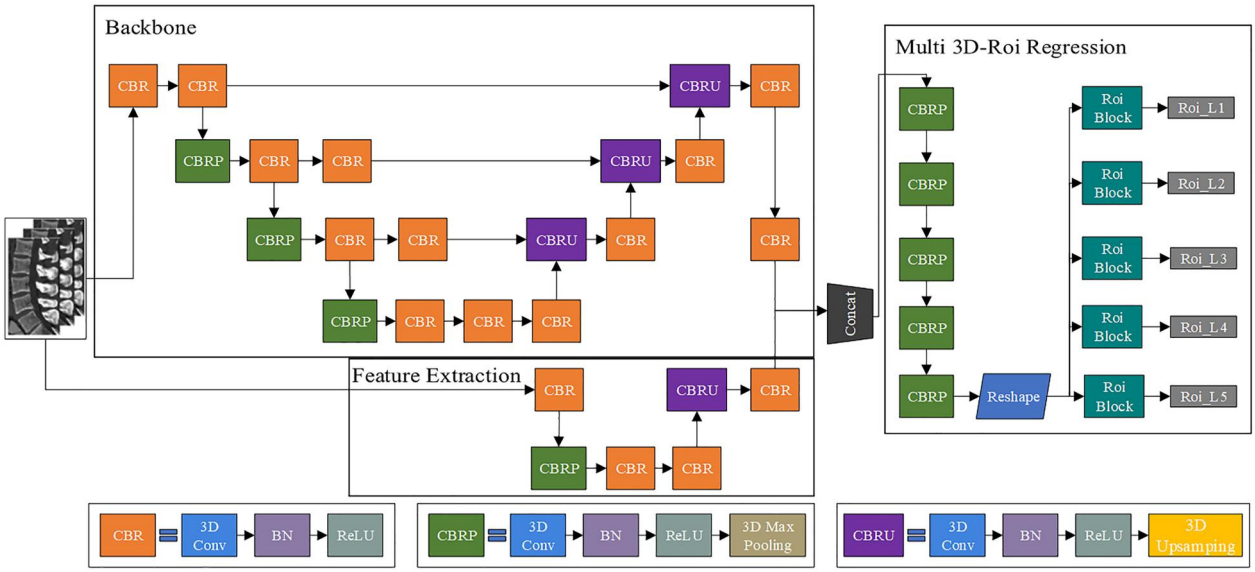

FIGURE 5 LMD-Net network structure diagram

feature fusion is realised by adding feature images output via the Feature Extraction and Backbone networks in a pixel-by-pixel manner.

Backbone in LMD-Net is mainly used to realise the sketchy spatial orientation of each lumbar spine. Inspired by Payer et al. [7, 18], we make use of Backbone to do heatmap regression for the lumbar landmark (the central point of the vertebral canal close to the cone), and L2 Loss is used to minimise the difference between the heatmap predicted by network and the target heatmap. As Backbone provides 'inspiration' for spatial orientation of lumbar spine, it is necessary to pre-train the Backbone network before the overall training of LMD-Net. The feature map of the local image processed by the Backbone network is shown in Figure 6.

The feature map output by the Backbone network is a feature map with stronger features in the region of the lumbar landmark, which expresses accurate spatial orientation information of the lumbar spine, but it is weak in expressing the size information of the lumbar spine. Therefore, it is necessary to extract the lumbar spine size information in local CT images through the Feature Extraction network. This network is relatively shallow and has a strong ability to extract details; therefore, it is able to extract information about the size of the spine by obtaining information about the edges of the spine. On adding the feature map output using the Backbone network and Feature Extraction network in a pixel-by-pixel manner, the fusion of the position feature and size feature can be realised. Subsequently, a fusion feature image can be obtained, which can effectively realise the spatial orientation of ROI_{L1}, ROI_{L2}…ROI_{L5}.

2.3 | 3D-MseCIOULoss

In LRP-Net and LMD-Net, the spatial location is achieved by non-linear regression. Inspired by complete-IoU Loss [20]

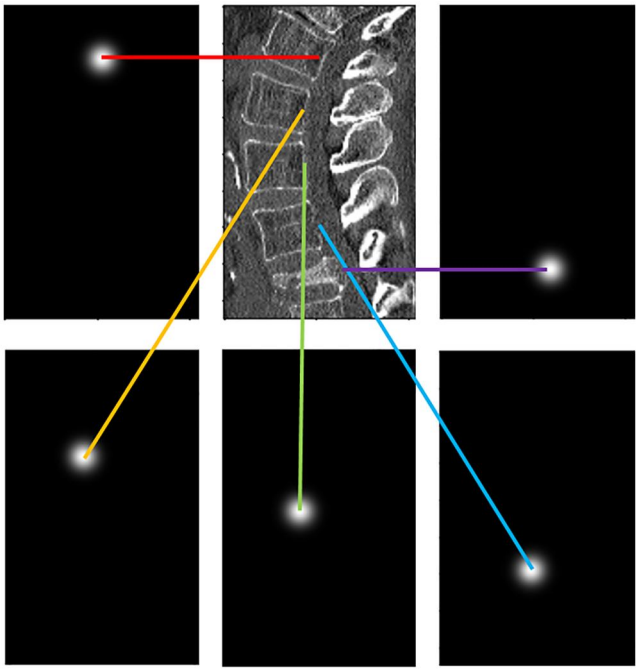

FIGURE 6 Feature map of the lumbar by Backbone of LMD-Net, there are five channels in total, and the light spot of each channel represents the general location of each lumbar landmark

(which is improved on the basis of IoU Loss [21]), we extend it to three-dimensional space to obtain 3D_CIOULoss. The process calculation method is as follows

$$\begin{aligned} 3D\_CIOULoss &= 1 - 3D\_CIOU = 1 - 3D\_DIOU \\ &+ \frac{\zeta^2}{1 - 3D\_IOU + \zeta}, \end{aligned} \tag{1}$$

where

$$3D\_DIOU = 3D\_IOU - \frac{Dis^2_{Center}}{Dis^2_{Corner}}, \tag{2}$$

$$\zeta = \frac{4}{\pi^2}\left[\arctan\left(\frac{ZDis_{Pre}}{XDis_{Pre}}\right) - \arctan\left(\frac{ZDis_{Tar}}{XDis_{Tar}}\right)\right]^2 + \frac{4}{\pi^2}\left[\arctan\left(\frac{YDis_{Pre}}{XDis_{Pre}}\right) - \arctan\left(\frac{YDis_{Tar}}{XDis_{Tar}}\right)\right]^2 \tag{3}$$

$$3D\_IOU = \frac{V_{Inter}}{V_{Pre}+V_{Tar}-V_{Inter}}, \tag{4}$$

where, V_{Inter} is the volume of the intersection of the space region cuboid ROI_{Pre} predicted by the network and the real target space region cuboid ROI_{Tar}. V_{Pre} is the volume of ROI_{Pre}, and V_{Tar} is the volume of ROI_{Tar}. Some parameters of the abovementioned formula (such as Dis_{Cen} and $ZDis_{Pre}$) are marked in Figure 7. 3D_CIOULoss is generally very effective; however, when the network is not initialised correctly, 3D_CIOULoss fails to effectively optimise the network. However, MSELoss does not require high random initialisation of the network; however, it does not consider the correlation between coordinates. Hence, its supervised learning effect on the network is inferior to that of 3D_CIOULoss. Thus, we combine the advantages of both and design 3D_MseCIOU-Loss. The calculation method is as follows:

$$3D\_MseCIOULoss = \partial \times MSELoss + \beta \times 3D\_CIOULoss \tag{5}$$

MSELoss in 3D_MseCIOULoss can effectively optimise the network during the initial stages of network training. When the network is optimised to a certain extent, 3D_CIOULoss can make full use of information, such as the coordinates, shapes and centre distance between ROI_Pre and ROI_Tar and realise effective network learning optimisation. The effects of MSELoss and 3D_CIOULoss on network training can be adjusted by varying the weight coefficients α and β. In general, $\partial < \beta$ and ∂ is very small.

3 | NETWORK TRAINING DETAILS

Because the orientation format of CT images should be the same, we redirected all CT images to RAI. As LRP-Net and LMD-Net both contain full connection layers, the size of the input image is fixed. The input image being resampled by LRP-Net is 224 × 112 × 112 (dimension order according to ITK: (z,y,x)), and the input image being resampled by LMD-Net is 192 × 96 × 96, all resampling processes are given cubic spline interpolation function. Whereas the Ground Truth Box of lumbar is automatically obtained by the label of image segmentation corresponding to CT images in the Verse2020 dataset [22–24]. The Backbone network of LMD-Net needs to be pre-trained by heatmap regression of the lumbar Landmark to achieve the purpose design of network. The Verse2020 dataset contains multiple Landmarks of CT images. We convert all coordinates to the RAI format with Spacing = (1,1,1); based on this, the target heatmap is generated. We use MSELoss to supervise and reduce the difference between the heatmap generated by LMD-Net prediction and the target heatmap. The calculation method for generating the target thermal map is as follows:

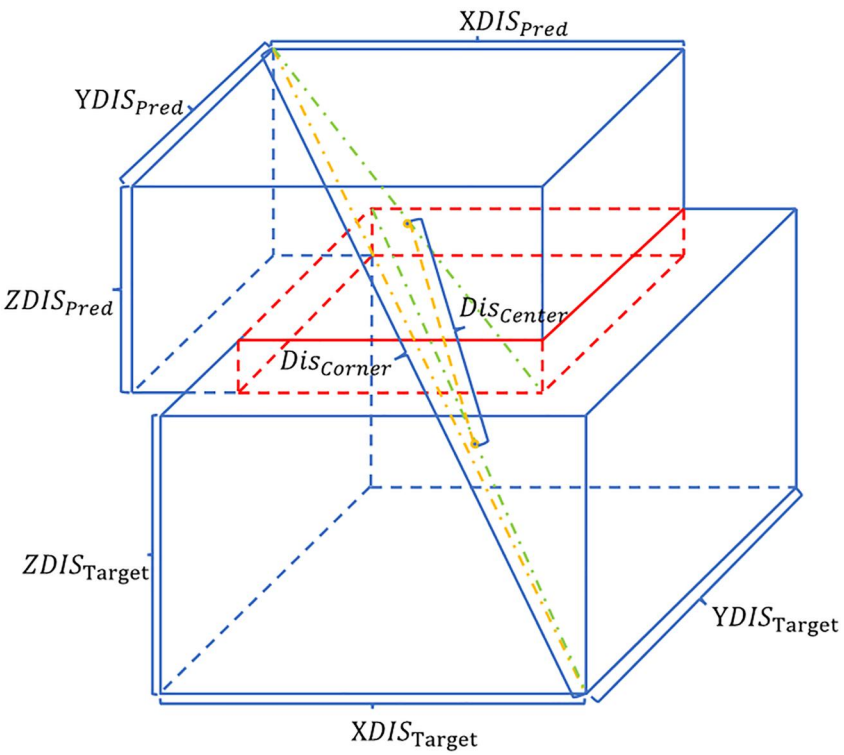

FIGURE 7 ROI_{Pre} and ROI_{Tar} can be represented by two cuboids. The proposed algorithm aims to maximise the 3D_IOU between ROI_{Pre} and ROI_{Tar}

$$G_i(x,\sigma) = \frac{1}{(2\pi)^{3/2}\sigma^3}\exp\left(-\frac{\|x-\dot{x}_i\|_2^2}{2\sigma^2}\right), \quad (6)$$

where $G_i(x,\sigma)$ represents the target heatmap of the ith channel, and $\dot{x}_i$ represents the coordinates of the ith Landmark. The function of Backbone network in LMD-Net provides 'inspiration' for spatial orientation of ROI_{L1}, ROI_{L2}…ROI_{L5}. Although the detection accuracy of Landmark is reduced on setting a larger value for σ, the probability of ambiguity is reduced considerably [18]. Therefore, we set $\sigma = 5$ during the pre-training of the Backbone of LMD-Net to generate the target heatmap. In addition, we improve the contrast between the spine and soft tissue by adjusting the window width and position of the CT images. This calculation method can be expressed as follows:

$$P_{Des} = \begin{cases} 0 & P_{Src} \le Min_{Bound} \\ \dfrac{P_{Src}-(Min_{Bound})}{Max_{Bound}-(Min_{Bound})} & Min_{Bound} < P_{Src} < Max_{Bound} \\ 1 & P_{Src} \ge Max_{Bound} \end{cases} \quad (7)$$

where $Min_{Bound} = -200$ and $Max_{Bound} = 600$. As shown in Figure 8, adjustment of window width and window position can effectively improve the contrast between the spine and the surrounding background.

Due to the limited number of CT images in the data set, we enhance the random date in the training process. We perform random grey scale transformation, random elastic transformation, and random rotation on the image. As LRP-Net and LMD-Net are both processing 3D CT images, the operations in them are all 3D operation. Parameters related to the convolution operation in LRP-Net and LMD-Net are set as kernel_size = 3 × 3 × 3, stride = 1 × 1 × 1, padding = 1 × 1 × 1. This ensures that the size of the image will not change due to the convolution operation. All pooling operations in the network are Max Pooling, and all parameters are set to kernel_size = 2 × 2 × 2, stride = 2 × 2 × 2. The size of each dimension of the image halves after this pooling operation. The upsampling operation parameters in the network (except the fine-tuning upsampling in CUS module) are scale_factor = 2 and mode = 'trilinear'. In the network, the activation function of the Backbone network output of LRP-Net and LMD-Net is Softmax, and the others are ReLU. To train the network more stably, we added a BatchNormal layer after each convolution operation [25]. In LMD-Net, the spatial orientation of ROI_{L1}, ROI_{L2}…ROI_{L5} is inspired by the Backbone network using heatmap regression; hence, the number of output feature channels for the Backbone and the Feature Extraction network of LMD-Net is set to five. The feature map of each channel is responsible for probing the spatial region of a lumbar spine.

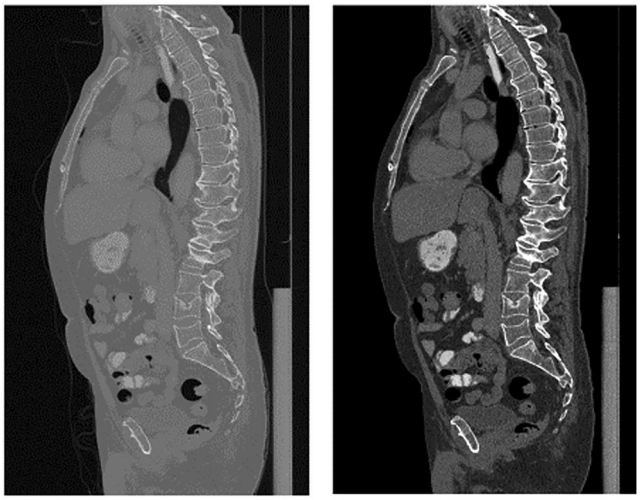

FIGURE 8 Window width and window position adjustment

For training, we set the weight coefficient in 3D_MseCIOULoss ($\partial = 0.01$ and $\beta = 1$). Through experimentation, it can be found that 3D_MseCIOULoss can achieve excellent performance under this setting.

During the training, 250 groups of CT data including lumbar spine from the Verse2020 dataset were randomly divided into training and verification sets, with 80% used for the training set and 20% used for the verification set. Our network training and verification were conducted on a server equipped with four NVIDIA Quadro RTX6000 graphics cards. We called two of them, and trained was performed for a total of 200 rounds; each round was also verified. The 200 rounds of LRP-Net training required 83 h, whereas the 200 rounds of LMD-Net training required 17 h.

4 | EXPERIMENT AND RESULTS

We tested our algorithm using 76 CT images from the Verse2020 dataset containing the lumbar region that did not participate in the training process as a test set. We conducted a number of experiments to test our method, including: ①Effect of increasing channel number on LRP-Net and its positioning accuracy; ②Comparison between LRP-Net and other algorithms; ③Effect of increasing the channel number on LMD-Net and its positioning accuracy; ④Comparison between LMD-Net and other algorithms; ⑤Comparison of LRP-Net and LMD-Net with other algorithms in operation speed.

We used Yolov3 [14], Faster RCNN [11], CenterNet [26] and the organ location algorithm for CT images proposed by Xu et al. [17] as our comparison algorithms to test the spatial localisation ability of LRP-Net and LMD-Net for the lumbar spine. These comparison algorithms were also trained and tested on the Verse2020 dataset. Moreover, according to the characteristics of each algorithm, we adjusted the corresponding hyper-parameters to achieve better training results. The corresponding algorithms obtained by the 3D extension were named 3D-Yolo, 3D-FasterRCNN, 3D-CenterNet and method of Xu et al [17].

4.1 | Evaluation indicators

The objective of this task is to detect the five lumbar spatial regions ROI_{L1}, ROI_{L2}…ROI_{L5}, that can be named directly from their respective spatial position. Therefore, there is no need to classify them through the network algorithms. Hence, we use 3D_IOU, Dis_{Cen} (distance between ROI_{Pre} centre coordinates and ROI_{Tar} centre coordinates) and SSP (Shape similarity parameter between ROI_{Pre} and ROI_{Tar}) to comprehensively evaluate the network accuracy. Among them, 3D_IOU is the main evaluation indicator of the network accuracy, and $Dis_{Cen}$and SSP are supplementary evaluation indicators. To verify the speed advantage of our algorithm compared with other algorithms, the average positioning time of the spine in each group of CT data in the test set was used as an evaluation indicator to evaluate the operation speed of our algorithm.

3D_IOU can describe the overlap between ROI_{Pre} and ROI_{Tar}, and its calculation method has been listed. As shown in Figure 9, Dis_{Cen} can describe the distance between the centre of ROI_{Pre} and ROI_{Tar}, particularly in the case of the same 3D_IOU parameter, the smaller the Dis_{Cen}, the better the ROI_{Pre}. We used Euclidean distance to calculate Dis_{Cen}. As shown in Figure 10, SSP can effectively describe the similarity of shapes between ROI_{Pre} and ROI_{Tar}, and $SSP \in (0, 1]$. The larger the SSP, the greater the similarity between ROI_{Pre} and ROI_{Tar} in terms of the shape, and vice versa. When 3D_IOU and Dis_{Cen} are the same, the larger the SSP, the better the detection effect of ROI_{Pre}. The SSP is calculated as

$$SSP = \frac{1}{1 + R}, \tag{8}$$

where

$$R = \left| \frac{ZDis_{Pre}}{XDis_{Pre}} - \frac{ZDis_{Tar}}{XDis_{Tar}} \right| + \left| \frac{YDis_{Pre}}{XDis_{Pre}} - \frac{YDis_{Tar}}{XDis_{Tar}} \right| \tag{9}$$

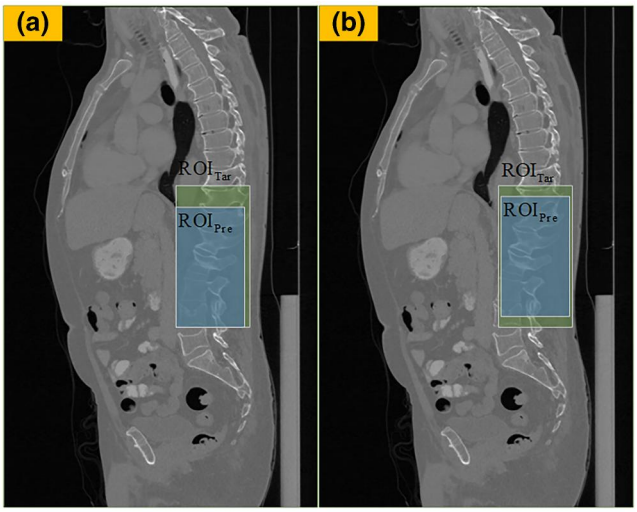

FIGURE 9 3D_IOU between ROI_{Pre} and ROI_{Tar} in group (a) and group (b) is the same, but the Dis_{Cen} of group (b) is smaller, so the effect of group (b) is better than that of group (a)

4.2 | LRP-Net test results

In the LRP-Net Backbone network, we used a fixed number of feature channels, that is, 32. We tested the effect of the channel number setting strategy on LRP-Net. Assuming the model size is roughly the same for both strategies of changing and fixing channel numbers, we set the change process for the feature channel number in each stage of the encoder as 8→ 16→ 32→ 64. The localisation effects of LRP-Net on ROI_{Lumbar} under the two strategies were tested. We calculated the 3D_IOU, Dis_{Cen} and SSP between ROI^{Pre}_{Lumbar} predicted by LRP-Net under both strategies and the real lumbar region ROI^{Tar}_{Lumbar}. We then obtained a number of statistics for these indicators, including the mean, median, standard deviation, maximum and minimum. These statistical results are listed in Table 1.

It can be seen from Table 1 that LRP-Net with a fixed number of feature channels has a higher positioning accuracy

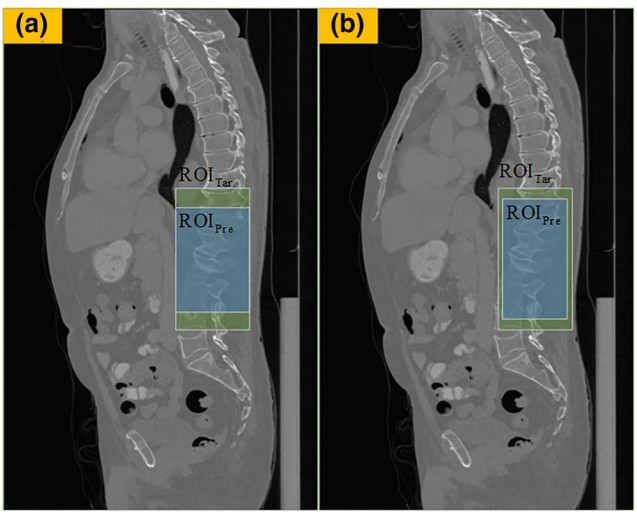

FIGURE 10 3D_IOU and Dis_{Cen} between ROI_{Pre} and ROI_{Tar} in group (a) and group (b) are the same, but SSP of group (b) is larger, so the effect of group (b) is better than that of group (a)

TABLE 1 Positioning results of ROI_{Lumbar} by LRP-Net under the two strategies

Strategies	Indicators	Mean	Med	Std	Max	Min
Increase	3D_IOU	0.8327	0.8573	0.1127	0.9856	0.5237
	Dis_{Cen}	2.7814	2.1474	2.0851	9.6217	0.5221
	SSP	0.8514	0.8819	0.1158	0.9698	0.4872
No increase	3D_IOU	0.8339	0.8598	0.0990	0.9881	0.5324
	Dis_{Cen}	2.6746	2.0616	2.0825	9.5131	0.5000
	SSP	0.8530	0.8841	0.1054	0.9731	0.5014

for ROI_{Lumbar} when the model sizes of the two strategies are the same. Therefore, a fixed number of feature channels is adopted in LRP-Net.

To visualise the distribution of the three evaluation indicators of LRP-Net on the positioning effect of ROI_{Lumbar}, we drew the corresponding distribution histogram, as illustrated in Figure 11.

From the above statistics, it is evident that LRP-Net has a concentrated distribution of the detecting effect on ROI_{Lumbar}, with 3D_IOU in the range of (0.7,0.9), Dis_{Cen} in the range of (0.5,3.9) and SSP in the range of (0.8,1.0). It is evident that LRP-Net is both accurate and robust.

Finally, the positioning effect of LRP-Net on the whole lumbar region is illustrated in Figure 12; the figure in the upper left side of the box depicts the 3D_IOU. It is evident that LRP-Net localised well for ROI_{Lumbar} in CT images with high noise and surgery history.

4.3 | Comparison of experimental results between LRP-Net and other algorithms

To demonstrate the effectiveness of LRP-Net, we used the four comparison algorithms selected above to conduct

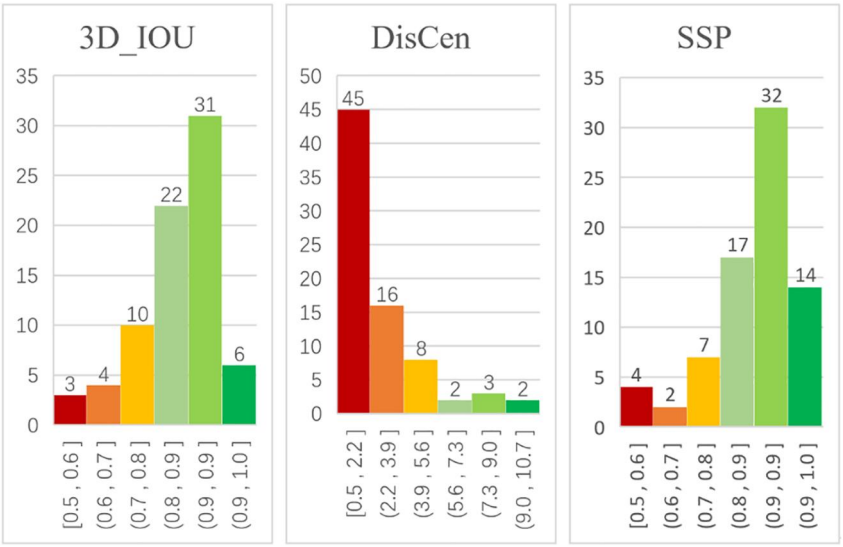

FIGURE 11 Distribution of LRP-Net detection accuracy

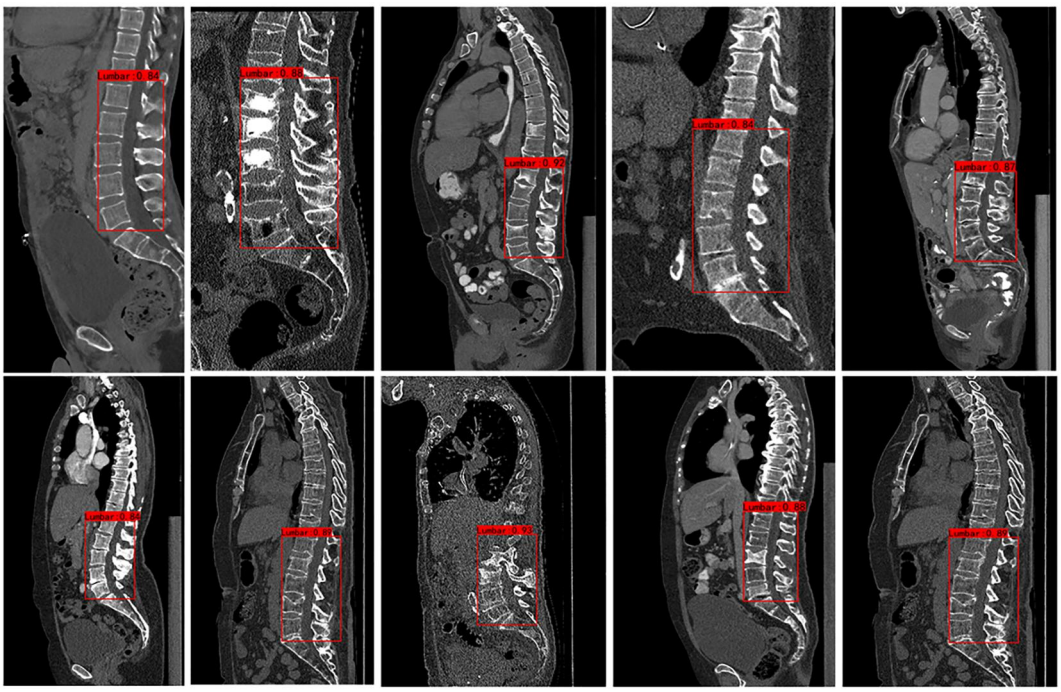

FIGURE 12 Positioning effect of LRP-Net

comparison experiments with LRP-Net. The positioning effect of each algorithm for ROI_{Lumbar} is shown in Table 2.

As shown in Table 2, compared with other algorithms, LRP-Net has a greater advantage in the localisation effect of ROI_{Lumbar}, and because it can be clearly shown from 3D_IOU, we did not calculate Dis_{Cen} and SSP extra.

4.4 | LMD-Net test results

Similar to LRP-Net, we used a fixed number of feature channels in the Backbone network of LMD-Net and set it to 48. To verify the effectiveness of our method, we tested the effect of increasing the number of feature channels on LMD-Net. Assuming that the model size is roughly the same under the two strategies of changing and fixing channel numbers, we set the change process of the feature channel number in each stage of the encoder as 8→ 16→ 32→ 64. The localisation effects of LMD-Net on ROI_{L1}, ROI_{L2}…ROI_{L5} under the two strategies were tested. We adopted the average level of LMD-Net's positioning effect for ROI_{L1}, ROI_{L2}…ROI_{L5} under the two strategies as the evaluation object, calculated three positioning accuracy evaluation indicators, and obtained the corresponding statistics. The average level is the average value of 3D_IOU located by LMD-Net on ROI_{L1}, ROI_{L2}…ROI_{L5} in the same group of CT. The test results are shown in Table 3.

It can be seen from Table 3 that increasing the number of feature channels has little impact on the effect of LMD-Net, and may even lead to a slight decline in performance. Therefore, we chose to adopt a fixed number of feature channels in the Backbone network.

To highlight the positioning effect of LMD-Net on ROI_{L1}, ROI_{L2}…ROI_{L5} in greater detail, we calculated three positioning accuracy indicators of LMD-Net for five lumbar vertebra segments and obtained their statistics. The experimental results are listed in Table 4.

To visualise the difference between the detection effects of LMD-Net on ROI_{L1}, ROI_{L2}…ROI_{L5}, the distribution of detection effect evaluation parameters is plotted in Figure 13.

It is evident from Figure 13 that the 3D_IOU is mainly distributed between (0.8,1.0), Dis_{Cen} is mainly distributed between (0.5,2.9) and SSP is mainly distributed between (0.8,1.0). In addition, LMD-Net has roughly the same detection ability for each lumbar spine from this distribution index, and there is no evident shortcoming.

The final detection effect of LMD-Net on each lumbar spine is shown in Figure 14, where the image in the upper left side of the box depicts the 3D_IOU.

4.5 | Comparison of experimental results between LMD-Net and other algorithms

Similar to LRP-Net, we used four selected comparison algorithms to conduct comparison experiments with LMD-Net, and the experimental results are shown in Table 5.

TABLE 2 3D_IOU statistics of ROI_{Lumbar} by LRP-Net and other algorithms()

Methods	Mean	Med	Std	Max	Min
3D-Yolo	0.7683	0.8063	0.1578	0.9519	0.5185
3D-FasterRCNN	0.8280	0.8529	0.1127	0.9587	**0.5483**
Method of Xu et al [17].	0.7860	0.7886	**0.0707**	0.9254	0.4150
3D-CenterNet	0.7185	0.7246	0.0721	0.8932	0.5020
LRP-Net	**0.8339**	**0.8598**	0.0990	**0.9881**	0.5324

Note: In the comparative experiment, the best performance of the indicators was processed in bold, so as to visually show the effect of each algorithm.

TABLE 3 Positioning results of ROI_{L1}, ROI_{L2}…ROI_{L5} by LMD-Net under the two strategies

Strategies	Indicators	Mean	Med	Std	Max	Min
Increase	3D_IOU	0.8540	0.8552	0.0361	0.8958	0.6389
	Dis_{Cen}	3.0159	2.8531	0.9095	8.0179	1.4362
	SSP	0.8820	0.8903	0.0374	0.9570	0.7605
No increase	3D_IOU	0.8559	0.8611	0.0332	0.9078	0.7441
	Dis_{Cen}	2.1527	2.0552	0.5084	3.4929	1.0699
	SSP	0.8988	0.9056	0.0316	0.9485	0.7938

TABLE 4 LMD-Net localisation results for five lumbar vertebrae

Objects	Indicators	Mean	Med	Std	Max	Min
L_1	3D_IOU	0.8621	0.8727	0.0592	0.9474	0.5727
	Dis_{Cen}	2.2360	2.1213	0.9372	4.3875	0.5000
	SSP	0.9213	0.9339	0.0480	0.9839	0.7233
L_2	3D_IOU	0.8462	0.8509	0.0510	0.9346	0.6832
	Dis_{Cen}	2.2665	2.1794	0.9164	4.3012	0.5000
	SSP	0.8971	0.9044	0.0520	0.9716	0.7247
L_3	3D_IOU	0.8572	0.8632	0.0458	0.9373	0.7281
	Dis_{Cen}	2.3443	2.2913	0.9972	5.2440	0.5000
	SSP	0.9086	0.9144	0.0520	0.9828	0.7070
L_4	3D_IOU	0.8703	0.8771	0.0447	0.9494	0.7544
	Dis_{Cen}	1.9250	1.8028	0.8320	4.5277	0.5000
	SSP	0.9049	0.9146	0.0490	0.9733	0.7215
L_5	3D_IOU	0.8507	0.8639	0.0583	0.9402	0.6212
	Dis_{Cen}	2.2156	2.0616	0.9330	4.5000	0.5000
	SSP	0.8747	0.8904	0.0748	0.9750	0.5580
Average	3D_IOU	0.8559	0.8611	0.0332	0.9078	0.7441
	Dis_{Cen}	2.1527	2.0552	0.5084	3.4929	1.0699
	SSP	0.8988	0.9056	0.0316	0.9485	0.7938

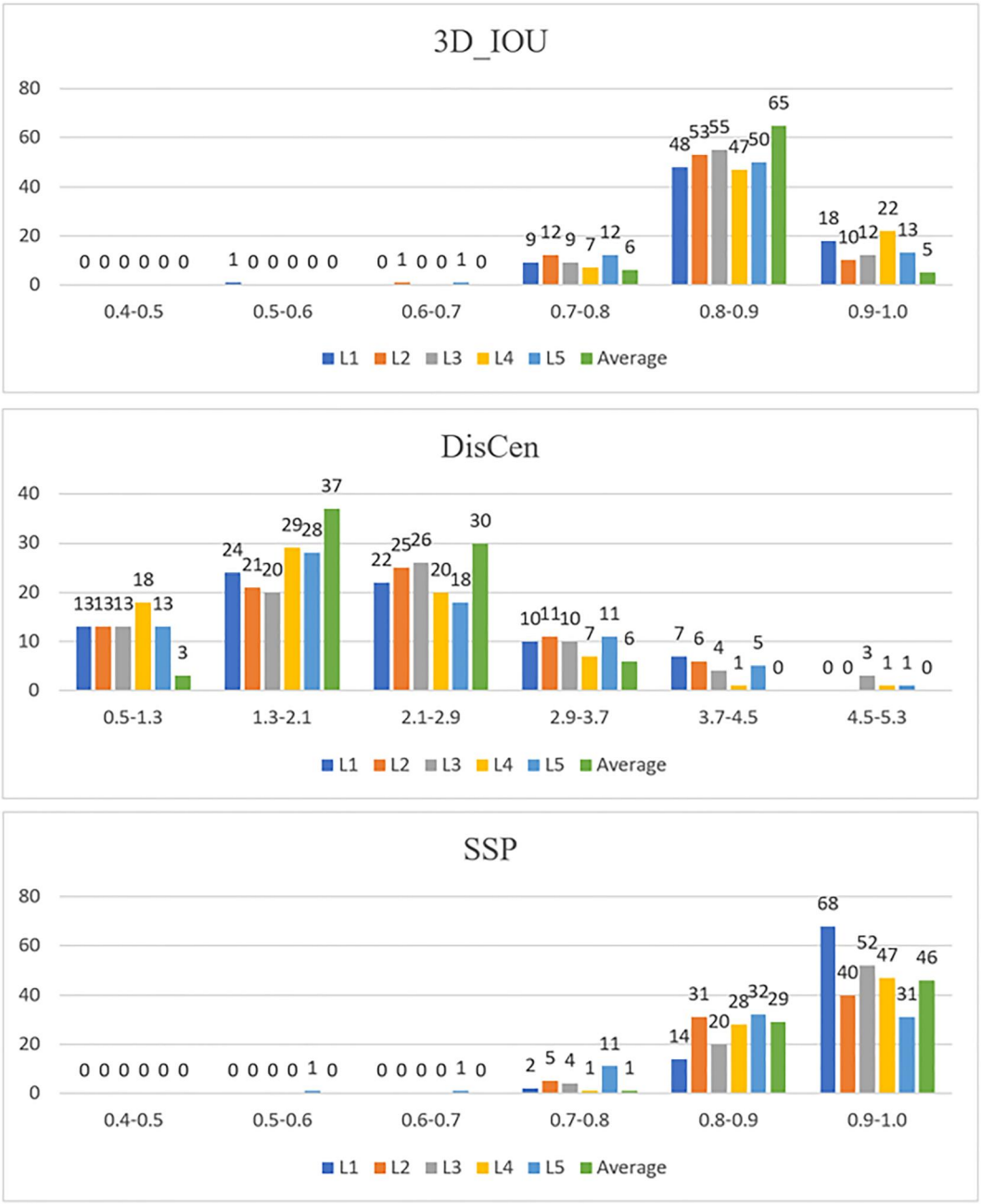

FIGURE 13 LMD-Net detection accuracy distribution of ROI_{L1}, ROI_{L2}…ROI_{L5}

As shown in Table 5, compared with other algorithms, LRP-Net has a greater advantage in the localisation effect of ROI_{L1}, ROI_{L2}…ROI_{L5}, and because it can be clearly shown from 3D_IOU, we did not calculate Dis_{Cen} and SSP extra.

In addition, to visualise the advantages and disadvantages of the detection ability of each comparison algorithm for ROI_{L1}, ROI_{L2}…ROI_{L5}, we plot the 3D_IOU distribution of each algorithm in Figure 15.

From the above figure, it is evident that LMD-Net is very useful for ROI_{L1}, ROI_{L2}…ROI_{L5}, the detection result is not only higher in accuracy, but also more concentrated in distribution, indicating that our algorithm is more robust.

4.6 | Comparison experiment of algorithm operation speed

Because most current target detection algorithms generate many candidate boxes, it is necessary to obtain the final target space region through non-maximum suppression or the weighting operation. However, the proposed LRP-Net and LMD-Net directly obtains the target region through non-linear regression, and the algorithm structure is relatively simple. Therefore, our algorithm offers significant advantages in terms of the operation speed. To verify the speed advantage of LRP-Net and LMD-Net, we conducted a comparative experiment

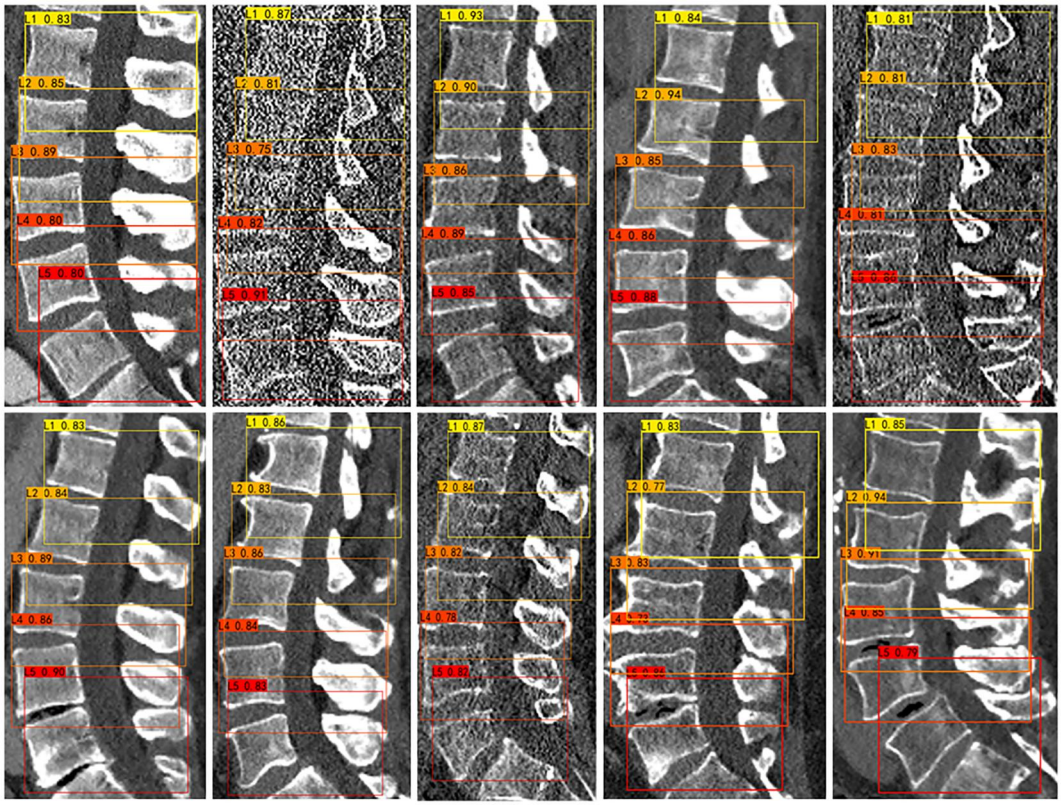

FIGURE 14 Detection effect of LMD-Net on each lumbar spine

TABLE 5 3D_IOU statistics of ROI_{L1}, ROI_{L2}…ROI_{L5} by LMD-Net and other algorithms

Methods	Objects	Mean	Med	Std	Max	Min
3D-Yolo	L_1	0.7513	0.7627	0.0905	0.9084	0.5405
	L_2	0.7392	0.7485	0.1153	**0.9351**	0.2234
	L_3	0.7667	0.7865	0.0787	0.9318	0.5422
	L_4	0.7707	0.7820	0.0761	0.8910	0.5435
	L_5	0.7546	0.7624	0.0871	0.9150	0.5183
	Average	0.7565	0.7562	0.0489	0.8552	0.5405
3D-FasterRCNN	L_1	0.7450	0.7582	0.0781	0.8921	0.5155
	L_2	0.7554	0.7500	0.0842	0.8944	0.5686
	L_3	0.7638	0.7904	0.0959	0.8984	0.5078
	L_4	0.7495	0.7569	0.0848	0.8975	0.6002
	L_5	0.7648	0.7804	0.1024	0.8985	0.2267
	Average	0.7557	0.7601	0.0479	0.8553	0.5317
Method of Xu et al.	L_1	0.7809	0.7918	0.0600	0.8716	**0.6041**
	L_2	0.7835	0.7862	0.0663	0.8962	0.5514
	L_3	0.8022	0.8130	0.0632	0.8998	0.6810
	L_4	0.8021	0.8197	0.0670	0.8955	0.6064
	L_5	0.7829	0.7944	0.0655	0.8857	0.5926
	Average	0.7903	0.7927	**0.0331**	0.8391	0.6134

TABLE 5 (Continued)

Methods	Objects	Mean	Med	Std	Max	Min
3D-CenterNet	L_1	0.7380	0.7478	0.0984	0.9074	0.4354
	L_2	0.7168	0.7345	0.1272	0.9549	0.2976
	L_3	0.7280	0.7541	0.1195	**0.9404**	0.4711
	L_4	0.7455	0.7576	0.0953	0.9239	0.4641
	L_5	0.6885	0.6914	0.1104	0.8682	0.4397
	Average	0.7234	0.7224	0.0529	0.8440	0.5453
LMD-Net	L_1	**0.8621**	**0.8727**	**0.0592**	**0.9474**	0.5727
	L_2	**0.8462**	**0.8509**	**0.0510**	0.9346	**0.6832**
	L_3	**0.8572**	**0.8632**	**0.0458**	0.9373	**0.7281**
	L_4	**0.8703**	**0.8771**	**0.0447**	**0.9494**	**0.7544**
	L_5	**0.8507**	**0.8639**	**0.0583**	**0.9402**	**0.6212**
	Average	**0.8559**	**0.8611**	**0.0332**	**0.9078**	**0.7441**

Note: In the comparative experiment, the best performance of the indicators was processed in bold, so as to visually show the effect of each algorithm.

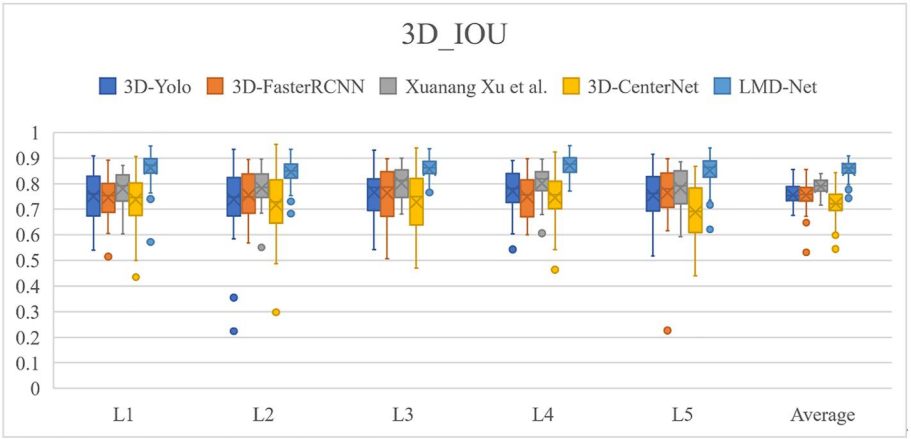

FIGURE 15 3D_IOU box diagram of detection results of ROI_{L1}, ROI_{L2}…ROI_{L5} by LMD-Net and other algorithms

TABLE 6 The time spent by each algorithm to achieve spinal localisation (the unit is in seconds)

Tasks	3D-Yolo	3D-FasterRCNN	Method of Xu et al.	3D-CenterNet	Ours
Positioning of ROI_{Lumber}	0.5329	0.9378	0.5705	0.4740	**0.3274**
Positioning of ROI_{L1}, ROI_{L2}…ROI_{L5}	0.3978	0.7068	0.5489	0.3903	**0.2105**

Note: In the comparative experiment, the best performance of the indicators was processed in bold, so as to visually show the effect of each algorithm.

of algorithm speed. We calculated the average time spent by each algorithm to locate the spine in each group of CT images in the test set, and the results were shown in Table 6.

As can be seen from Table 6, the proposed LRP-Net and LMD-Net have obvious advantages in computing speed compared with other algorithms.

5 | DISCUSSION

The proposed two-stage method for spatial orientation of each lumbar vertebra in CT images shows good accuracy and high computational efficiency. Although our method shows better results in both accuracy and speed than the most advanced

methods in the lumbar precise positioning task, there are still some limitations.

First, we have not achieved single-stage direct localisation of each lumbar segment in the original CT image, and as a result our method did not run at maximum efficiency. In addition, both LRP-Net and LMD-Net are robust to lumbar spatial localisation on spinal CT images in the presence of high noise, spinal deformity, compression fracture, and bone cement. However, its lumbar spatial positioning accuracy with metal artefacts left by surgery is a little low. Finally, due to the high dimension of CT images and the three-dimensional operation of the algorithm, the BatchSize cannot be set too high during training, at the same time the hardware requirements are very high.

Therefore, in the future, we will devote ourselves to studying a network algorithm with clever structure, which can not only achieve the precise positioning of the single-stage lumbar space region, but also effectively simplify our network structure and decrease the hardware requirements. In addition, we will enhance the robustness of our algorithm by enriching the data set and adding more random data enhancement methods during training.

6 | CONCLUSION

In this study, to provide a good basis for lumbar image segmentation and surgical path planning, we proposed a two-stage lumbar spatial region localisation method. In addition, two algorithms were proposed, LRP-Net and LMD-Net, which show a high level of accuracy and speed. The 3D_IOU of LRP-Net and LMD-Net can reach 0.8339 ± 0.0990 and 0.8559 ± 0.0332, and the average time of localisation is only 0.3274 and 0.2105 s respectively. Compared with other target detection algorithms, our algorithm has obvious advantages. In addition, 3D-MseCIOUloss is proposed, which can effectively supervise the iterative training of LRP-Net and LMD-Net. In the future, we will focus on designing a more ingenious network structure to achieve single-stage lumbar spatial region localisation and also reduce the complexity of the network structure; this is expected to improve the accuracy and efficiency of the algorithm. In addition, we plan to enrich the data set and optimise the network training process to improve the robustness of this algorithm.

ACKNOWLEDGEMENTS

The authors thank all anonymous reviewers for their thorough work and excellent comments that helped us to improve our manuscript. Moreover, the authors also thank the MICCAI 2020 organisers for making the Verse2020 dataset freely available. This work was supported by the Beijing Natural Science Funds-Haidian Original Innovation Joint Fund: L202010 and the National Key Research and Development Program of China: 2018YFB1307604. National Key Research and Development Program of China, Grant/Award Numbers:2018YFB1307604

CONFLICT OF INTEREST

The author declares that they have no conflict of interest.

DATA AVAILABILITY STATEMENT

The data that support the findings of this study are available in verse2020 at https://github.com/anjany/verse.

ORCID

Yonghong Zhang https://orcid.org/0000-0002-0622-8503

REFERENCES

1. Katz, J.N., Harris, M.B.: Lumbar spinal stenosis. N. Engl. J. Med. 358(8), 818–825 (2008). https://doi.org/10.1056/nejmcp0708097
2. Deyo, R.A., et al.: United States trends in lumbar fusion surgery for degenerative conditions. Spine 30(12), 1441–1445 (2005). https://doi.org/10.1097/01.brs.0000166503.37969.8a
3. Li, Q., Du, Z., Yu, H.: Trajectory planning for robot-assisted laminectomy decompression based on Ct images. IOP Conf. Ser. Mater. Sci. Eng. 768(4), 042037 (2020). https://doi.org/10.1088/1757-899x/768/4/042037
4. Sun, Y., et al.: Robot-assisted decompressive laminectomy planning based on 3D medical image. IEEE Access 6, 22557–22569 (2018). https://doi.org/10.1109/access.2018.2828641
5. Janssens, R., Zeng, G., Zheng, G.: IEEE, fully automatic segmentation of lumbar vertebrae from Ct images using cascaded 3D fully convolutional networks. In: 15th IEEE International Symposium on Biomedical Imaging (ISBI) (2018)
6. Sekuboyina, A., et al.: A Localisation-Segmentation Approach for Multi-Label Annotation of Lumbar Vertebrae Using Deep Nets. arXiv e-prints (2017)
7. Payer, C., et al.: Coarse to fine vertebrae localization and segmentation with spatial configuration-Net and U-Net. In: 15th International Joint Conference on Computer Vision, Imaging and Computer Graphics Theory and Applications (VISIGRAPP)/15th International Conference on Computer Vision Theory and Applications (VISAPP) (2020)
8. Lessmann, N., et al.: Iterative fully convolutional neural networks for automatic vertebra segmentation and identification. Med. Image Anal. 53, 142–155 (2019). https://doi.org/10.1016/j.media.2019.02.005
9. Girshick, R., et al.: IEEE, rich feature hierarchies for accurate object detection and semantic segmentation. In: 27th IEEE Conference on Computer Vision and Pattern Recognition (CVPR) (2014)
10. Girshick, R.: IEEE, fast R-Cnn. In: IEEE International Conference on Computer Vision (2015)
11. Ren, S., et al.: Towards real-time object detection with region proposal networks. IEEE Trans. Pattern Anal. Mach. Intell. 39(6), 1137–1149 (2017). https://doi.org/10.1109/tpami.2016.2577031
12. Redmon, J., et al.: IEEE, You only look once: unified, real-time object detection. In: 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). IEEE (2016)
13. Redmon, J., Farhadi, A.: IEEE, 'Yolo9000: better, faster, stronger'. In: 30th IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) (2017)
14. Redmon, J., Farhadi, A.: Yolov3: An Incremental Improvement. arXiv e-prints (2018)
15. Bochkovskiy, A., Wang, C.Y., Liao, H.: Yolov4: Optimal Speed and Accuracy of Object Detection. arXiv e-prints (2004)
16. de Vos, B.D., et al.: 2D image classification for 3D anatomy localization: employing deep convolutional neural networks. In: Conference on Medical Imaging - Image Processing (2016)
17. Xu, X., et al.: Efficient multiple organ localization in Ct image using 3D region proposal network. IEEE Trans. Med. Imag. 38(8), 1885–1898 (2019). https://doi.org/10.1109/tmi.2019.2894854
18. Payer, C., et al.: Integrating spatial configuration into Heatmap regression based Cnns for landmark localization. Med. Image Anal. 54, 207–219 (2019). https://doi.org/10.1016/j.media.2019.03.007

19. Ronneberger, O., Fischer, P., Brox, T.: U-Net: convolutional networks for biomedical image segmentation. In: 18th International Conference on Medical Image Computing and Computer-Assisted Intervention (MICCAI) (2015)
20. Zheng, Z., et al.: Assoc advancement artificial, I., distance-IOU loss: faster and better learning for bounding box regression. In: 34th AAAI Conference on Artificial Intelligence/32nd Innovative Applications of Artificial Intelligence Conference/10th AAAI Symposium on Educational Advances in Artificial Intelligence (2020)
21. Yu, J., et al.: Unit box: An Advanced Object Detection Network. ACM (2016)
22. Sekuboyina, A.: Verse: A Vertebrae Labelling and Segmentation Benchmark. arXiv e-prints (2020)
23. Löffler, M.T., et al.: A vertebral segmentation dataset with fracture grading, radiology. Artif. Intell. 2(4), e190138 (2020). https://doi.org/10.1148/ryai.2020190138
24. Liebl, H., et al.: A computed tomography vertebral segmentation dataset with anatomical variations and multi-vendor scanner data. Sci. Data 8(1), 284 (2021). https://doi.org/10.1038/s41597-021-01060-0
25. Ioffe, S., Szegedy, C.: Batch normalization: accelerating deep network training by reducing internal covariate shift. In: 32nd International Conference on Machine Learning (2015)
26. Zhou, X., Wang, D., Krähenbühl, P.: Objects as Points. arXiv e-prints (2019)

How to cite this article: Zhang, Y., et al.: Lumbar spine localisation method based on feature fusion. CAAI Trans. Intell. Technol. 1–15 (2022). https://doi.org/10.1049/cit2.12137