

MBNet: A Multi-Task Deep Neural Network for Semantic Segmentation and Lumbar Vertebra Inspection on X-ray Images

Van Luan Tran¹[0000-0001-8060-3036], Huei-Yung Lin¹[0000-0002-6476-6625], and Hsiao-Wei Liu²[0000-0002-1627-8117]

¹ Department of Electrical Engineering,

National Chung Cheng University, Chiayi 621, Taiwan

² Biomedical Technology and Device Research Laboratories,

Industrial Technology Research Institute, Hsinchu 310, Taiwan

tranvanluan07118@gmail.com, lin@ee.ccu.edu.tw, rachel_liu@itri.org.tw

Abstract. Deep learning methods, especially multi-task learning with CNNs, have achieved good results in many fields of computer vision. Semantic segmentation and shape detection of lumbar vertebrae, sacrum, and femoral heads from clinical X-ray images are important and challenging tasks. In this paper, we propose a multi-task deep neural network, MBNet. It is developed based on our new multi-path convolutional neural network, BiLuNet, for semantic segmentation on X-ray images. Our MBNet has two branches, one is for semantic segmentation of lumbar vertebrae, sacrum, and femoral heads. It shares the main features with the second branch to learn and classify by supervised learning. The output of the second branch is to predict the inspected values for lumbar vertebra inspection. These networks are capable of performing the two tasks with very limited training data. We collected our dataset and annotated it by doctors for model training and performance evaluation. Compared to the state-of-the-art methods, our BiLuNet model provides better mIoUs with the same training data. The experimental results have demonstrated the feasibility of our MBNet for semantic segmentation of lumbar vertebrae, as well as the parameter prediction for the doctors to perform clinical diagnosis of low back pains. Code is available at: <https://github.com/LuanTran07/BiLuNet-Lumbar-Spine>.

1 Introduction

Semantic segmentation and shape detection are fundamental to the object identification and location computation in medical images [1, 2]. While the X-ray inspection is of great clinical necessity, the automatic understanding of X-ray images is still a very challenging task. It is primarily due to the projective nature of X-ray imaging, which causes large overlaps among the object shapes, blurred boundaries, and complicated composition patterns [3]. For specific medical applications such as the treatment for low back pain, the images of spine bones are acquired by X-ray [4], computed tomography [5], or magnetic resonance imaging [6, 7], to facilitate the assessment of the causes by doctors. Consequently,

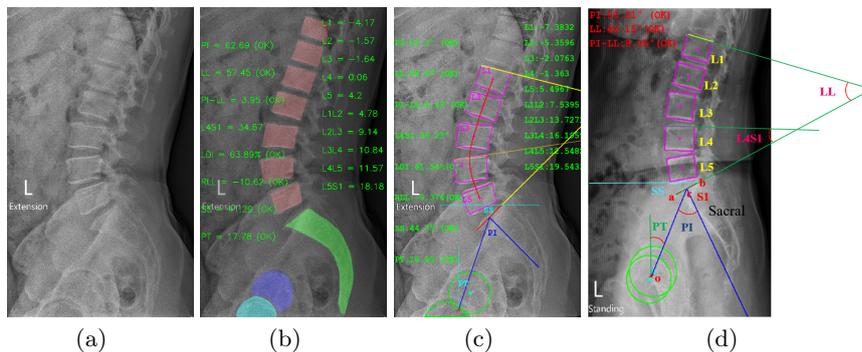


Fig. 1. An overview of the lumbar vertebra inspection on X-ray images by our multi-task deep neural network for semantic segmentation and inspection. (a) the input X-ray image. (b) our MBNet generates the semantic segmentation result and inspected values. The vertebrae, sacrum, and femoral heads are segmented with red, green, blue, and cyan color respectively. (c) annotation reference is provided by doctors. (d) defined for the parameters used for lumbar vertebra inspection, S1 is defined by the line segment ab constructed between the posterior superior corner of the sacrum and the anterior tip of the S1 endplate. The hip axis (o) is the midpoint of the line segment connecting the centers of femoral heads. The pelvic radius is drawn from the hip axis to the center of S1.

the automatic detection and segmentation of vertebrae in medical images is an important step for the analysis of the disease severity and recovery from the treatment.

The recent state-of-the-art techniques for vertebra detection, segmentation, and localization from images usually adopt two main strategies: the approaches using traditional computer vision and image processing algorithms and the methods based on deep neural networks. The former utilizes image enhancement, feature extraction, clustering, and classification for local region segmentation [8–10], and the latter adopts convolutional neural networks with the models trained using the datasets labeled by the experts in the field [11, 12]. Due to the overlapping shadows of chest organs, including lungs, bowels, and other bony structures such as ribs, the automatic detection, and segmentation of vertebrae using traditional methods are more difficult for X-ray images than for CT or MRI images [3, 13, 14]. Nevertheless, the use of X-ray imaging techniques is popular because of the least amount of radiation.

With the traditional approach, Ebrahimi *et al.* proposed a technique for posterior corner detection in X-ray images of the lumbar spine [9]. They first enhanced the image region of the lumbar area and performed Canny edge detection. The corner points were then selected via Haar-filtering to automatically extract the positions of lumbar (L) vertebrae posterior corners for L1 to L5. The results provided a localization accuracy at about 0.6 ~ 1.4 mm, which was comparable to the manual expert variability. In [15], Larhman *et al.* presented

a method for vertebra detection on X-ray images using the generalized Hough transform (GHT). They adopted contrast-limited adaptive histogram equalization (CLAHE) to enhance the image for edge detection. A GHT accumulator construction was then used to determine the position of the center points of vertebrae in the X-ray image with the information stored in the table created based on the vertebra model. Their detection results demonstrated an 89% global accuracy from 200 investigated vertebrae.

In the past few years, the progress of deep learning with convolutional neural networks (CNNs) has shown a remarkable performance on biomedical image segmentation. However, the existing literature for CNN based lumbar vertebra segmentation is mostly related to the techniques which take MRI images as inputs and is very limited for the processing of X-ray images [16]. Fully automatic segmentation on X-ray images for the lumbar vertebrae still has many challenges. In [12], Arif *et al.* proposed a novel shape predictor network for vertebra segmentation with X-ray images. They modified the U-Net [17] architecture to generate a signed distance function (SDF) from the input image. The predicted SDF was then converted to the parameters represented in a shape space for loss computation. After the shape was transferred to the corresponding symbolic distance function, a principal component analysis (PCA) was performed. However, their results were still very limited since it was not easy to derive the shape information for complicated cases.

As machine learning techniques become more powerful, multi-task deep neural networks also gain considerable attention [10, 18–21]. Multi-task learning (MTL) performs model training for several tasks simultaneously to share information in the network and solve multiple tasks at the same time. In [22], Liu *et al.* proposed a multi-task deep learning model for automated lung nodule analysis. They built a bi-branch model used for the classification and attribute score regression tasks. This method makes the neural network more reliable in detecting marginal lung nodules. In [23], Xiao *et al.* proposed a new regularized multi-task learning model to jointly select a small set of standard features across multiple tasks. The experimental results demonstrated that the model achieved high performance compared with other methods.

In this paper, we present a multi-task network for the semantic segmentation of lumbar vertebrae as well as the parameter prediction for inspection and diagnosis. It is designed based on our effective semantic segmentation network, BiLuNet, for the robust detection of unclear overlapping shadow regions in the images. For comparison, BiLuNet is trained independently to determine the location of five lumbar (L1 to L5), S1 on the sacrum and the hip axis. The parameters for the diagnosis of lumbar vertebrae are then derived. In the experiments, the training data are labeled at the pixel level manually and annotated by doctors. The datasets are used to train our MBNet and BiLuNet, as well as other networks for performance comparison. Experimental results have demonstrated the effectiveness and feasibility of our proposed MBNet for both the semantic segmentation and lumbar vertebra inspection on X-ray images with very high accuracy.

1.1 Background

The inspection of the lumbar vertebrae includes object segmentation, target localization, and analysis of several inclination angles in the X-ray images. Figure 1(d) depicts the surgical measurements of interest in this work for identification and future examination. Pelvic incidence (PI) is defined as the angle between the line perpendicular to the sacral plate at its midpoint (c) and the line connecting this point to the axis of femoral heads [24]. PI is an important pelvic anatomic parameter which determines the pelvic orientation. It is a specific constant for each individual and remains relatively stable during childhood. Afterward, it increases significantly during youth until reaching its maximum, and does not change after adolescence [25].

Pelvic tilt (PT) is defined as the angle between the anterior pelvic plane (APP) and a vertical line in the standing position. To be more specific, it is the angle given by the line passing through the hip axis o and the line segment \overline{oc} in the X-ray image (see Fig. 1(d)). Sacral slope (SS) is defined as the angle between the sacral endplate S1 and the horizontal axis passing through its endpoint (b). PT and SS are position-dependent angles in the spatial orientation of the pelvis [26]. PI, SS, and PT are particularly useful since it can be shown that PI is the arithmetic sum of SS and PT, and these two position-dependent variables can determine the pelvic orientation in the sagittal plane [27].

Lumbar lordosis (LL) is the inward curve of the lumbar spine, and defined as the angle between the lines drawn along the superior endplate of L1 and the superior endplate of S1 [28]. The angle L4S1 is measured by the intersection of the lines drawn along the superior endplate of L4 and the superior endplate of S1. Most bodyweight and movement are carried by the two lowest segments in the lumbar spine, L5/S1, and L4/L5, which makes the area prone to injury. In 95% of disc herniated cases, L4/L5 or L5/S1 levels are involved. Low back pain with degenerative disc disease is when normal changes take place in the discs of the spine, mostly at the sites L4/L5 lumbar vertebrae and S1 sacrum [29, 30]. The surgery for patients with unbalanced lumbar spondylolisthesis is helpful to correct the spine-pelvic sagittal parameters (PI, PT, SS, L4S1, and LL) closer to the normal range [31].

PI-LL is the value of PI minus LL. Given a PI value, it offers an estimate of the lordosis and quantifies the mismatch between pelvic morphology and the lumbar curve. Lordosis distribution index (LDI) defines the magnitude of the arc lordosis (L4S1) in proportion to the arc of the lumbar lordosis (LL). Relative lumbar lordosis (RLL) quantifies the magnitude of lumbar lordosis relative to the ideal lordosis as defined by the magnitude of PI. RLL and LDI are used to describe the state where the lumbar lordosis is rehabilitated to the ideal magnitude and distribution. The evaluation of RLL and PI-LL provides the prediction of postoperative complications and the correlation with health-related quality of life (HRQOL) scores [32].

2 Multi-Task Deep Neural Network (MBNet)

In this section, we describe our MBNet architecture, where M represents the multi-task network structure and B denotes the BiLuNet backbone model. Our MBNet automatically learns appropriate representation sharing through end-to-end training, and makes full use of the correlations among different tasks. The MBNet with supervised learning is applied for two tasks, semantic segmentation and parameter inspection for the diagnosis of lumbar vertebrae as presented in Figure 3.

2.1 Semantic Segmentation Branch

For semantic segmentation, this work is specifically concerned with an accurate segmentation method to extract the object mask. It is used to detect the locations of lumbar vertebrae, sacrum, and femoral heads. In this regard, U-Net is one of the well-known fully convolutional networks (FCNs) utilized in medical image understanding [17]. It has many variants in the literature, and most of them consist of an encoder-decoder structure. However, U-Net can hardly segment well for challenging cases or the image with a complex background. To deal with these issues, we propose a new multi-path convolutional neural network, BiLuNet, where Lu represents the shape of the network structure and Bi denotes the bilateral segmentation network. As shown in Figure 2, our network has two main paths, a spatial path to recover the spatial information and a downsampling path to extract the features.

In BiLuNet, Spatial Path is to preserve the spatial information and generate high-resolution features [33]. It contains three layers, and each convolutional layer has $stride = 2$, followed by batch normalization and Rectified Linear Unit (ReLU). Thus, this path extracts the output feature maps at 1/8 of the original image size. It encodes the rich spatial information from the large spatial size of the feature maps. The spatial path and downsampling path are crucial to the prediction of the detailed output to cover large objects, which leads to a rich discriminative ability.

The spatial information acquired by Spatial Path encodes most rich and detailed messages in the low levels, while the robust features extracted from the downsampling path are in the high levels. These two feature paths with different levels are fused by a Feature Fusion Module (FFM) [33]. Given the features from different levels, we first concatenate the outputs of the spatial path and the downsampling path. The batch normalization is then utilized to balance the scales of the features. Next, a global pool is used for the concatenated features to form a feature tensor, and the weight is computed by a sigmoid function. This weight tensor can re-weight the features by multiplying the concatenated features and adding to the feature selection. The details of this FFM function is shown in Figure 2. The dimension of the output space is 5, which corresponds to the 5 classes of the segmentation requirement (background, vertebra, sacrum, femoral head1, and femoral head2).

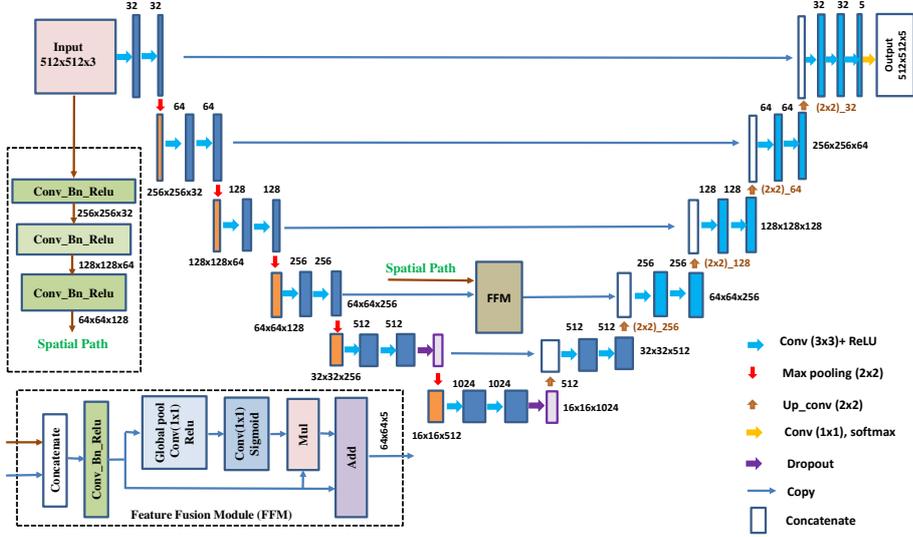


Fig. 2. The end-to-end BiLuNet architecture for object semantic segmentation. White boxes present the concatenation results of a copied feature map and an *up_conv*, an up-sampled and convolutional layer. Purple boxes indicate the results of the node dropped between two convolutional layers. BiLuNet takes a three-channel input image at the resolution of 512×512 , and the output contains five channels, the classes with background, vertebra, sacrum, femoral head1, and femoral head2.

Similar to U-Net, the proposed BiLuNet operates with both the downsampling and upsampling paths and adds a spatial path to recover the spatial information lost in the pooling layers of the contracting path. The downsampling path decreases the feature size while increasing the number of feature maps, whereas the upsampling path increases the feature size while decreasing the number of feature maps. The latter eventually leads to a pixel-wise mask. For the upsampling operation, we modify the existing architecture to reconstruct high-resolution feature maps. The detailed network structure with the parameter settings is illustrated in Figure 2.

To maintain the same number of channels in the encoder and decoder, skip connections are added between the corresponding blocks in the downsampling and upsampling paths. The skip connections are also used to explicitly copy features from the earlier layers to the subsequent layers. This prevents the neural network from having to learn the identity functions. Further functions are solely tasked with building on the representation. Through the skip connections, the nodes from a shallow layer are concatenated to the nodes in a deeper layer. The deeper layer is then treated as a wider layer and connected to the next layer. We use a dropout layer between these two convolutional layers to prevent overfitting and co-adaptation of the features. In the implementation, 50% of the nodes are

dropped out during the forward and backward passes. The network totally has 29 convolutional layers. In the final layer, a 1×1 convolution, 5 dimensional space, and a sigmoid activation function are used to output the probability map of the semantic segmentation, which has the same size as the original 512×512 input.³

2.2 Inspected Values Branch

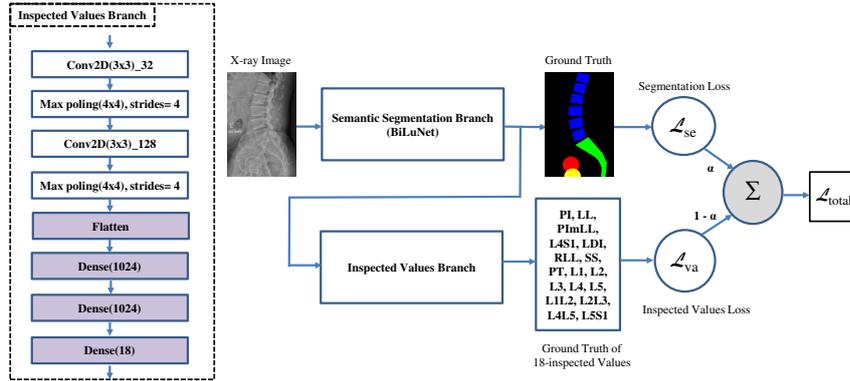


Fig. 3. MBNet architecture. It consists of two branches. The semantic segmentation branch is trained to produce a semantic segmentation of the lumbar vertebrae, sacrum, and femoral heads. The inspected values branch generates the inspected values for the related parameters for the diagnosis of the lumbar vertebrae.

Besides generating the semantic segmentation map, our network also has other functions, which can estimate the physical quantities. It is automatically estimated to provide accurate parameters for the doctors to perform clinical diagnosis of low back pains. Here, the classification task is considered as an auxiliary task. While the semantic segmentation task provides the local supervision for the segmentation of lumbar vertebrae, sacrum, and femoral heads from clinical X-ray images. The regression values level classification task provides global supervision about how inspected values. The integrated classification part can drive our model more robust to the inspected values variation and boost performance. Via the dataset, the segmentation mask, and the inspected values ground truth, we train our network to predict the parameters used for lumbar vertebra inspection.

As presented in Figure 3, the input of the inspected values branch is shared from the final layer of the semantic segmentation branch. The inspected values

³ The full training and testing code is open source at <https://github.com/LuanTran07/BiLUNet-Lumbar-Spine>.

branch designed to be extracted from the shared feature maps. This network branch consisting of two 3x3 convolutional layers, two 4x4 max-pooling layers, and three fully connected (FC) layers were constructed as an inspected values classifier. The first and second FC layer consists of 1024 neurons with the rectified linear unit activation function. The output FC layer consists of 18 neurons followed by a linear activation function, indicating the probabilities of the 18-inspected values. To train the inspected values classifier supervised learning, the mean squared error regression loss is used, which can be calculated by equation 3.

2.3 Multiple Loss Function

Our loss function is a weighted sum of a semantic segmentation loss (L_{se}) and an inspected values loss (L_{va}):

$$L_{total} = \alpha \cdot L_{se} + (1 - \alpha) \cdot L_{va} \quad (1)$$

where $\alpha \in [0, 1]$ is loss weights. It adjusts the relative importance of L_{se} and L_{va} . The loss weights (α) is set to 0.9 in our method.

We use categorical cross-entropy loss function for the semantic segmentation loss. The categorical cross-entropy is given by

$$L_{se} = -\frac{1}{N} \sum_{i=1}^N \sum_{c=1}^C y_c \cdot \log(\hat{y}_c) + (1 - y_c) \cdot \log(1 - \hat{y}_c) \quad (2)$$

where y_j and \hat{y}_j denote the predicted and ground-truth values of the object in the image, respectively. C is the number of categories, and N is the number of training samples. The categorical cross-entropy cost function works in tandem with a softmax activation function. It computes the loss between network predictions and target values for multi-class classification.

The inspected values loss is defined by mean squared error regression loss. It is the most commonly used regression loss function. The mean squared error (MSE) is the sum of squared distances between our target variable and predicted values. The inspected values loss is given by

$$L_{va} = \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{n} \quad (3)$$

where y_i and \hat{y}_i denote the predicted and ground-truth values of the inspected values in the X-ray image, respectively. n is the number of training samples.

3 Experiments

In this work, we collect our own image data for the experiments. The dataset consists of 650 images (with a half before surgery and half after surgery) for network training, and additional 100 images for testing (also with a half before

surgery and half after surgery). The training images are labeled with four categories (vertebra, sacrum, femoral head1, and femoral head2), ground-truth of 18-inspected values are annotated by doctors. For the CPU configuration, the computer is with an Intel i7-7700 CPU at 3.6 GHz and 32GB RAM. The GPU is an NVIDIA TITAN RTX with 24GB RAM. For training parameters, the number of epochs is set to 200, and the batch size is set to 4. Our initial learning rate is set to $3e - 4$ for Adam.

3.1 Semantic Segmentation Evaluation

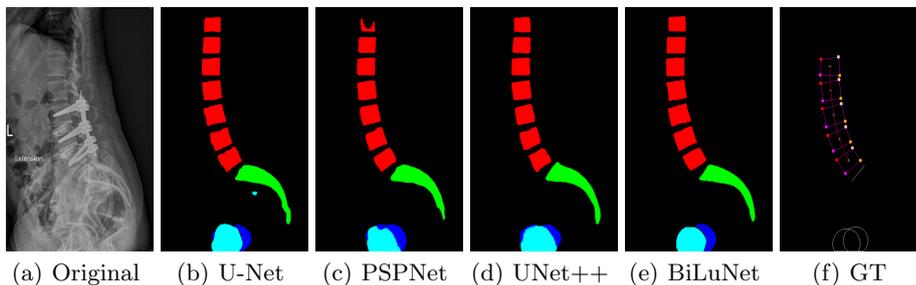


Fig. 4. An example of lumbar vertebra segmentation output on a challenging X-ray image. Our result shows a significant improvement compared to those derived from U-Net, UNet++, and PSPNet.

The testing dataset is also used to evaluate other methods for performance comparison with the same training parameters and data. Figure 5 illustrates the mIoU (mean intersection over union) of U-Net, UNet++, PSPNet, and our BiLuNet. In 100 testing images, the proposed network has 84.10%, 83.67%, 80.35%, 78.67% and 98.48% mIoUs for vertebrae, sacrum, femoral head1, femoral head2 and background, respectively. The proposed BiLuNet outperforms state-of-the-art methods in all categories. Figure 4 also shows the higher accuracy semantic segmentation results compared to others. Our model can cover the most region of input X-ray images by the feature fusion between the spatial path and the downsampling path. The spatial path designed in our network preserves the spatial information from the original X-ray image. As shown in the figure, very precise results of lumbar vertebrae, sacrum, and femoral heads are obtained. More specifically, we are able to derive the accurate shapes of femoral head1 and femoral head2, which are failed by all other methods.

We conduct several experiments to demonstrate the effectiveness of different modules in our framework. The ablation study is carried out with three cases. In the first case, the network is implemented without Spatial Path and Feature Fusion Module. The network in the second case does not include FFM, but use a concatenation function to combine Spatial Path and the downsampling

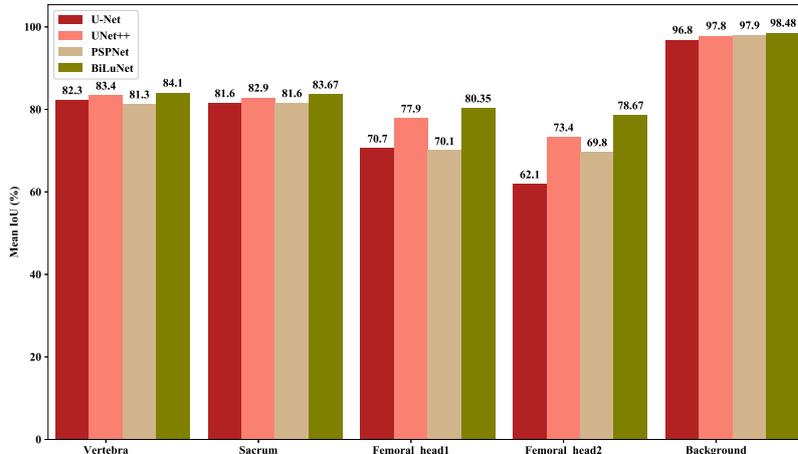


Fig. 5. The performance comparison on our dataset for different algorithms, U-Net [17], UNet++ [34], and PSPNet [35], and our BiLuNet. It shows the IoU (intersection over union) and mean IoU for the evaluation between the semantic segmentation testing results and ground-truth labels.

Table 1. Ablation study on different cases

Methods	Vertebrae	Sacrum	Femoral head1	Femoral head2	Background	Average mIoU
Case 1	82.98%	81.95%	72.89%	60.91%	95.90%	78.93%
Case 2	83.15%	82.71%	77.26%	69.42%	97.96%	80.64%
BiLuNet	84.10%	83.67%	80.35%	78.67%	98.48%	85.05%

path. The third case denotes the full BiLuNet. Table 1 reports the mIoUs (mean intersection over union) of the three cases using 100 images for evaluation. The average mIoUs are 78.93%, 82.10% and 85.05%, respectively. It demonstrates that the spatial path designed in our network does improve the segmentation results by preserving the spatial information from the X-ray image input. The ablation analysis also shows that our network can effectively combine two feature paths fused by Feature Fusion Module.

3.2 Lumbar Vertebra Inspection Results

The results of our MBNet tested on the images acquired before and after the operations are shown in Figures 6(a) – 6(c) and Figures 6(e) – 6(g), respectively. Two sample test X-ray images with the resolution of 2035×3408 captured before and after surgery are shown in Figures 6(a) and 6(e), respectively. The semantic segmentation is carried out with four target classes (vertebra, sacrum, femoral head1, and femoral head2) represented by red, green, blue and cyan, respectively (see Figs. 6(b) and 6(f)). As shown in Figures 6(b) and 6(f), the inspection

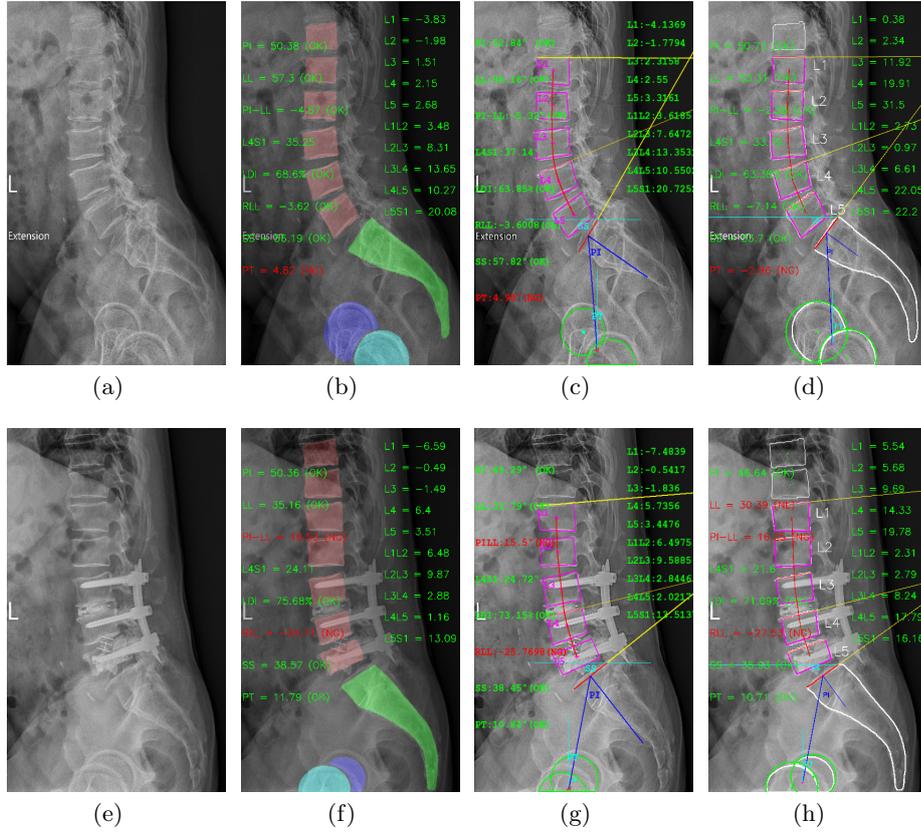


Fig. 6. The lumbar vertebra inspection with our MBNet and BiLuNet on X-ray images. (a), (e) the input X-ray images before and after surgery, respectively, (b), (f) Our MBNet results with the inspection results and the semantic segmentation of vertebrae, sacrum, left and right hip joints, in red, green, blue, and cyan, respectively, (c), (g) the reference annotation by a doctor, (d), (h) Our BiLuNet results with deriving the parameters from the segmentation and the identified shape and location of these objects (L1, L2, L3, L4, L5, S1, hip axis).

results of our MBNet are very close to the annotation provided by doctors (Figs. 6(c) and 6(g)). To the best of our knowledge, no related studies are published for comparison. Thus, BiLuNet is trained with the same parameters and data for semantic segmentation of vertebrae, sacrum, and femoral heads. The post-processing is then carried out to identify the shape and location of the objects (L1, L2, L3, L4, L5, S1, hip axis). The locations of five vertebrae (L1 to L5), sacrum (S1) and hip axis (o) are used to derive the parameters for lumbar vertebra inspection as shown in Figures 6(d) and 6(h).

Table 2 shows the standard ranges of common interests, PI, LL, PImLL, LDI, RLL, SS and PT [36]. The angles PI and PT are determined by the center of S1

and the hip axis, so they are affected by the locations of femoral heads and S1. More specifically, the derivation of most parameters depends on the accuracy of S1. PImLL (PI – LL) presents an appreciation of the lordosis required for a given PI value and quantifies the mismatch between pelvic morphology and the lumbar curve [32]. The value LDI and the angle RLL are defined by

$$LDI = (L4S1/LL) \quad (4)$$

$$RLL = LL - (0.62 \cdot PI + 29) \quad (5)$$

In the experiments, the inspection result is marked as OK if the value is in the standard range.

To evaluate our parameter estimation with respect to the annotated values (ground-truth), we define the accuracy as

$$Acc = 1 - \frac{|v_e - v_a|}{r_s} \quad (6)$$

where v_e and v_a are the parameter estimates of our inspection and the value annotated by doctors, respectively, and r_s is the standard range shown in Table 2. The accuracy evaluation of lumbar vertebra inspection is shown in Table 3 and 4. Comparative accuracy levels are obtained for both the preoperative and postoperative images.

Table 2. The standard range for our lumbar vertebra inspection.

Parameter	PI	LL	PImLL	LDI	RLL	SS	PT
Standard range	34°~84°	31°~79°	-10°~10°	50%~80%	-14°~11°	-20°~65°	5°~30°

Table 3. The accuracy evaluation for the testing image shown in Figure 6 (before surgery) for our MBNet and BiLuNet. Acc is the accuracy of our inspection results with respect to the annotation by doctors (GT).

Names	MBNet	BiLuNet	Ground-truth	Acc MBNet (%)	Acc BiLuNet (%)
PI°	50.38	50.73	52.84	95.08	95.78
LL°	57.30	53.31	58.16	98.21	89.90
PImLL°	-4.67	-2.58	-5.32	96.75	86.30
LDI%	68.60	63.38	63.85	84.17	98.43
RLL°	-3.62	-7.14	-3.60	99.92	85.84
SS°	56.19	53.70	57.82	98.08	95.15
PT°	4.82	-2.96	4.98	99.36	68.24

This testing set consists of 100 X-ray images, with a half before and a half after surgeries. The testing images are selected randomly based on the femoral

Table 4. The accuracy evaluation for the testing image shown in Figure 6 (after surgery) for our MBNet and BiLuNet. Acc is the accuracy of our inspection results with respect to the annotation by doctors (GT).

Names	MBNet	BiLuNet	Ground-truth	Acc MBNet(%)	Acc BiLuNet (%)
PI°	50.36	46.64	49.29	97.86	95.08
LL°	35.16	30.39	33.79	97.15	98.21
PImLL°	16.53	16.25	15.50	94.85	96.75
LDI%	75.68	71.09	73.15	91.57	84.17
RLL°	-24.71	-27.53	-25.77	95.76	99.92
SS°	38.57	35.93	38.45	99.86	98.08
PT°	11.79	10.71	10.84	96.2	99.36

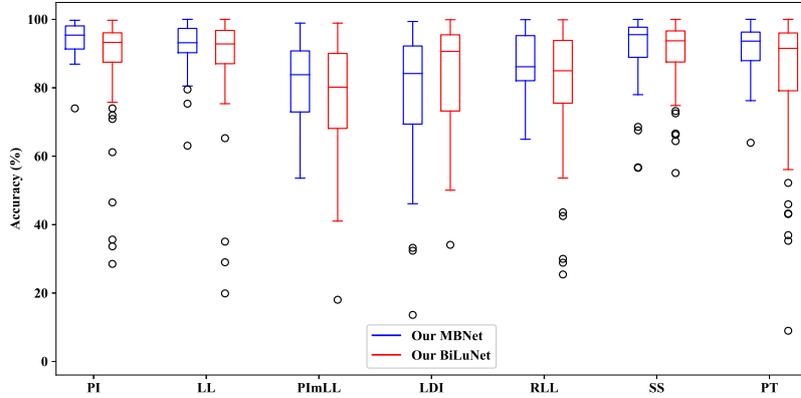


Fig. 7. The accuracy evaluation for Our MBNet and BiLuNet using 100 X-ray images (with a half before surgery and half after surgery). They are calculated based on the comparison with the inspection result annotated by doctors.

head locations. Figure 7 shows the evaluation results of 100 images with our MBNet and BiLuNet. MBNet is able to achieve the high average accuracy of PI, LL, PImLL, LDI, RLL, SS, and PT with 94.48%, 92.10%, 81.09%, 78.55%, 86.63%, 91.29%, and 91.15%, respectively. The mean accuracy of LDI is lower than other cases due to the limited training data for a large variation range. BiLuNet is able to achieve the average accuracy of PI, LL, PImLL, LDI, RLL, SS, and PT with 88.30%, 89.24%, 78.14%, 83.90%, 81.36%, 90.10%, and 83.77%, respectively. The parameter inspection of PT, PI, PImLL, and RLL is affected by the femoral head locations and less accurate as expected. The X-ray images with femoral heads unclear or missing are considered as challenging cases. One typical example is shown in Figure 8, where the second femoral head is not available for segmentation.

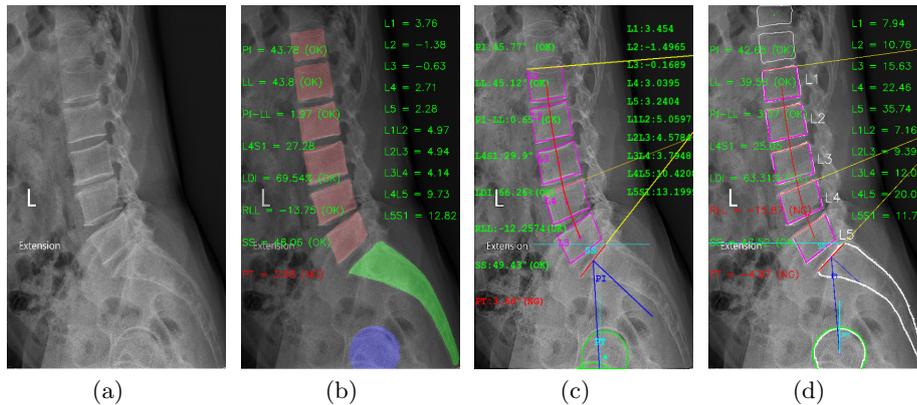


Fig. 8. Our MBNet and BiLuNet results with the challenging case of the second femoral head. (a) the input X-ray images before surgery, (b) Our MBNet results with the inspection results and the semantic segmentation of vertebrae, sacrum, left and right hip joints, in red, green, blue and cyan, respectively, (c) is the reference annotation by a doctor, (d) Our BiLuNet results with deriving the parameters from the segmentation and the identified shape and location of these objects (L1, L2, L3, L4, L5, S1, hip axis).

4 Conclusion

In this paper, we propose a multi-task deep neural network for the semantic segmentation of lumbar vertebrae as well as the parameter prediction for inspection and diagnosis. A multi-path convolutional neural network, BiLuNet, is presented for semantic segmentation on X-ray images. It is able to achieve high accuracy for the automated segmentation of lumbar vertebrae, sacrum, and femoral heads. In performance evaluation, our BiLuNet model with global pooling and spatial path outperforms other state-of-the-art semantic segmentation methods. Our MBNet is able to provide accurate parameters to the doctors for clinical diagnosis of low back pains. The experimental results have demonstrated that our approach is able to achieve high accuracy on lumbar vertebra inspection.

Acknowledgment

The authors would like to thank the support of this work in part by the Ministry of Science and Technology of Taiwan under Grant MOST 106-2221-E-194-004 is gratefully acknowledged.

References

1. Sekuboyina, A., Rempfler, M., Kukacka, J., Tetteh, G., Valentinič, A., Kirschke, J.S., Menze, B.H.: Btrfly net: Vertebrae labelling with energy-based adversarial learning of local spine prior. In: Medical Image Computing and Computer

- Assisted Intervention - MICCAI 2018 - 21st International Conference, Granada, Spain, September 16-20, 2018, Proceedings, Part IV. (2018) 649–657
2. Han, Z., Wei, B., Leung, S., Chung, J., Li, S.: Towards automatic report generation in spine radiology using weakly supervised framework. In: Medical Image Computing and Computer Assisted Intervention - MICCAI 2018 - 21st International Conference, Granada, Spain, September 16-20, 2018, Proceedings, Part IV. (2018) 185–193
 3. Zhang, Y., Miao, S., Mansi, T., Liao, R.: Task driven generative modeling for unsupervised domain adaptation: Application to x-ray image segmentation. In: Medical Image Computing and Computer Assisted Intervention - MICCAI 2018 - 21st International Conference, Granada, Spain, September 16-20, 2018, Proceedings, Part II. (2018) 599–607
 4. Horng, M., Kuok, C., Fu, M., Lin, C., Sun, Y.: Cobb angle measurement of spine from x-ray images using convolutional neural network. *Comp. Math. Methods in Medicine* **2019** (2019) 6357171:1–6357171:18
 5. Kalichman, L., Kim, D.H., Li, L., Guermazi, A., Hunter, D.J.: Computed tomography-evaluated features of spinal degeneration: prevalence, intercorrelation, and association with self-reported low back pain. *The spine journal : official journal of the North American Spine Society* **10 3** (2010) 200–8
 6. Jamaludin *et al.*, A.: Automation of reading of radiological features from magnetic resonance images (mris) of the lumbar spine without human intervention is comparable with an expert radiologist. *European Spine Journal* **26** (2017)
 7. Fu *et al.*, M.: Inter-rater and intra-rater agreement of magnetic resonance imaging findings in the lumbar spine: Significant variability across degenerative conditions. *The Spine Journal* **14** (2014)
 8. Tang, M., Valipour, S., Zhang, Z.V., Cobzas, D., Jägersand, M.: A deep level set method for image segmentation. In: Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support - Third International Workshop, DLMIA 2017, and 7th International Workshop, ML-CDS 2017, Held in Conjunction with MICCAI 2017, Québec City, QC, Canada, September 14, 2017, Proceedings. (2017) 126–134
 9. Ebrahimi, S., Angelini, E., Gajny, L., Skalli, W.: Lumbar spine posterior corner detection in x-rays using haar-based features. In: 2016 IEEE 13th International Symposium on Biomedical Imaging (ISBI). (2016) 180–183
 10. Moeskops, P., Wolterink, J.M., van der Velden, B.H.M., Gilhuijs, K.G.A., Leiner, T., Viergever, M.A., Isgum, I.: Deep learning for multi-task medical image segmentation in multiple modalities. *CoRR* **abs/1704.03379** (2017)
 11. Ebner, M., Wang, G., Li, W., Aertsen, M., Patel, P.A., Aughwane, R., Melbourne, A., Doel, T., David, A.L., Deprest, J., Ourselin, S., Vercauteren, T.: An automated localization, segmentation and reconstruction framework for fetal brain MRI. In: Medical Image Computing and Computer Assisted Intervention - MICCAI 2018 - 21st International Conference, Granada, Spain, September 16-20, 2018, Proceedings, Part I. (2018) 313–320
 12. Al-Arif, S.M.M.R., Knapp, K., Slabaugh, G.G.: Spnet: Shape prediction using a fully convolutional neural network. In: Medical Image Computing and Computer Assisted Intervention - MICCAI 2018 - 21st International Conference, Granada, Spain, September 16-20, 2018, Proceedings, Part I. (2018) 430–439
 13. Li, X., Dou, Q., Chen, H., Fu, C.W., Qi, X., Belavy, D., Armbrrecht, G., Felsenberg, D., Zheng, G., Heng, P.A.: 3d multi-scale fcn with random modality voxel dropout learning for intervertebral disc localization and segmentation from multi-modality mr images. *Medical image analysis* **45** (2018) 41–54

14. Liu, C., Zhao, L. In: Intervertebral Disc Segmentation and Localization from Multi-modality MR Images with 2.5D Multi-scale Fully Convolutional Network and Geometric Constraint Post-processing. (2019) 144–153
15. Larhmam, M.A., Mahmoudi, S., Benjelloun, M.: Semi-automatic detection of cervical vertebrae in x-ray images using generalized hough transform. In: 2012 3rd International Conference on Image Processing Theory, Tools and Applications (IPTA). (2012) 396–401
16. Nie, D., Gao, Y., Wang, L., Shen, D.: Asdnet: Attention based semi-supervised deep networks for medical image segmentation. In: Medical Image Computing and Computer Assisted Intervention - MICCAI 2018 - 21st International Conference, Granada, Spain, September 16-20, 2018, Proceedings, Part IV. (2018) 370–378
17. Ronneberger, O., Fischer, P., Brox, T.: U-net: Convolutional networks for biomedical image segmentation. In: Medical Image Computing and Computer-Assisted Intervention - MICCAI 2015 - 18th International Conference Munich, Germany, October 5 - 9, 2015, Proceedings, Part III. (2015) 234–241
18. Girard, N., Charpiat, G., Tarabalka, Y.: Aligning and updating cadaster maps with aerial images by multi-task, multi-resolution deep learning. In Jawahar, C.V., Li, H., Mori, G., Schindler, K., eds.: Computer Vision - ACCV 2018 - 14th Asian Conference on Computer Vision, Perth, Australia, December 2-6, 2018, Revised Selected Papers, Part V. Volume 11365 of Lecture Notes in Computer Science., Springer (2018) 675–690
19. Khaleel, A., Tasar, O., Charpiat, G., Tarabalka, Y.: Multi-task deep learning for satellite image pansharpening and segmentation. (2019) 4869–4872
20. Chen, Q., Peng, Y., Keenan, T.D., Dharssi, S., Agrón, E., Wong, W.T., Chew, E.Y., Lu, Z.: A multi-task deep learning model for the classification of age-related macular degeneration. CoRR [abs/1812.00422](https://arxiv.org/abs/1812.00422) (2018)
21. Li, X., Hu, X., Yu, L., Zhu, L., Fu, C.W., Heng, P.A.: Canet: Cross-disease attention network for joint diabetic retinopathy and diabetic macular edema grading (2019)
22. Liu, L., Dou, Q., Chen, H., Olatunji, I.E., Qin, J., Heng, P.: Mtmr-net: Multi-task deep learning with margin ranking loss for lung nodule analysis. In: Deep Learning in Medical Image Analysis - and - Multimodal Learning for Clinical Decision Support - 4th International Workshop, DLMIA 2018, and 8th International Workshop, ML-CDS 2018, Held in Conjunction with MICCAI 2018, Granada, Spain, September 20, 2018, Proceedings. Volume 11045 of Lecture Notes in Computer Science., Springer (2018) 74–82
23. Xiao, L., Stephen, J.M., Wilson, T.W., Calhoun, V.D., Wang, Y.: A manifold regularized multi-task learning model for IQ prediction from two fmri paradigms. IEEE Trans. Biomed. Engineering **67** (2020) 796–806
24. Sohn *et al.*, S.: Sagittal spinal alignment in asymptomatic patients over 30 years old in the korean population. Acta Neurochirurgica **159** (2017)
25. Labelle, H., Roussouly, P., Berthonnaud, E., Dimnet, J., O’Brien, M.: The importance of spino-pelvic balance in l5-s1 developmental spondylolisthesis. Spine **30** (2005) S27–34
26. Janssen, M., Kouwenhoven, J.W., Castelein, R.: The role of posteriorly directed shear loads acting on a pre-rotated growing spine: A hypothesis on the pathogenesis of idiopathic scoliosis. Studies in health technology and informatics **158** (2010) 112–7
27. Weinberg, D., Morris, W., Gebhart, J., Liu, R.: Pelvic incidence: an anatomic investigation of 880 cadaveric specimens. European Spine Journal **25** (2015)
28. Singh, R., Yadav, S., Sood, S., Yadav, R., Rohilla, R.: Spino-pelvic radiological parameters in normal indian population. SICOT-J **4** (2018) 14

29. Natarajan, R., Andersson, G.: Lumbar disc degeneration is an equally important risk factor as lumbar fusion for causing adjacent segment disc disease: Fusion, degeneration adjacent disc disease. *Journal of Orthopaedic Research* **35** (2016)
30. Legaye, J. In: *Analysis of the Dynamic Sagittal Balance of the Lumbo-Pelvi-Femoral Complex*. (2011)
31. Kim, M., Lee, S.H., Kim, E.S., Eoh, W., Chung, S.S., Lee, C.S.: The impact of sagittal balance on clinical results after posterior interbody fusion for patients with degenerative spondylolisthesis: A pilot study. *BMC musculoskeletal disorders* **12** (2011) 69
32. Yilgor *et al.*, C.: Relative lumbar lordosis and lordosis distribution index: individualized pelvic incidence-based proportional parameters that quantify lumbar lordosis more precisely than the concept of pelvic incidence minus lumbar lordosis. *Neurosurgical focus* **43** (2017) E5
33. Yu, C., Wang, J., Peng, C., Gao, C., Yu, G., Sang, N.: Bisenet: Bilateral segmentation network for real-time semantic segmentation. *CoRR* **abs/1808.00897** (2018)
34. Zhou, Z., Siddiquee, M.M.R., Tajbakhsh, N., Liang, J.: Unet++: A nested u-net architecture for medical image segmentation. *CoRR* **abs/1807.10165** (2018)
35. Zhao, H., Shi, J., Qi, X., Wang, X., Jia, J.: Pyramid scene parsing network. In: *2017 IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2017, Honolulu, HI, USA, July 21-26, 2017*. (2017) 6230–6239
36. Zheng *et al.*, G.Q.: Relationship between postoperative lordosis distribution index and adjacent segment disease following l4-s1 posterior lumbar interbody fusion. *Journal of Orthopaedic Surgery and Research* **15** (2020)