Accurate Segmentation of Lumbar Spine Magnetic Resonance Imaging by Deep Neural Network with Size-Adaptive Loss

YINGDI, ZHANG*
Key Laboratory of Opto-Electronic Information Processing, Chinese Academy of Sciences, Shenyang, China, Shenyang Institute of Automation, Chinese Academy of Sciences, Shenyang, China, Institutes for Robotics and Intelligent Manufacturing, Chinese Academy of Sciences, Shenyang, China, University of Chinese Academy of Sciences, Beijing, China

ZELIN, Shi
Shenyang Institute of Automation, Chinese Academy of Sciences, Shenyang, China

HUAN, Wang
Spine Surgery Department, ShengJing Hospital of China Medical University, Shenyang, China

CHONGNAN, Yan
Spine Surgery Department, ShengJing Hospital of China Medical University, Shenyang, China

LANBO, Wang
Spine Surgery Department, ShengJing Hospital of China Medical University, Shenyang, China

YUEMING, Mu
Spine Surgery Department, ShengJing Hospital of China Medical University, Shenyang, China

YUNPENG, Liu
Shenyang Institute of Automation, Chinese Academy of Sciences, Shenyang, China

SHUHANG, Wu
Shenyang Institute of Automation, Chinese Academy of Sciences, Shenyang, China

TIANCI, Liu
Shenyang Institute of Automation, Chinese Academy of Sciences, Shenyang, China

MINGQI, Pang
Shenyang Institute of Automation, Chinese Academy of Sciences, Shenyang, China

ABSTRACT

Accurate segmentation of lumbar spine is crucial for the assessment of spine disease diagnosis. The accuracy and robustness of segmentation is challenged by morphological complexity of lumbar spine anatomy structure, low contrast and background noise. Moreover, unbalanced anatomical structure impede the segmentation of small-size anatomy. To overcome these problems, an automatic segmentation method of lumber spine MRI is needed. In this paper, we propose to segment lumbar spine using a convolutional neural network (CNN) model. Based on the canonical u-shaped model, a multi-scale feature fusion module (MSFF) is proposed as bottleneck to better capture lumbar spine anatomy information of different sizes. To address the problem of anatomy size imbalance, we reshape the standard cross entropy loss as Size-Adaptive loss such that it adaptively weights the loss assigned to anatomy with different sizes. In order to fully verify the accuracy and effectiveness of the proposed method, we construct the first lumbar spine anatomy segmentation dataset which consists of 747 lumbar axial images from 83 patients. The dice similarity coefficient (DSC) is used to evaluate the segmentation methods. Compared with other methods, the proposed method achieves the best results with the highest average DSC of 0.8911. Ablation study shows that the proposed MSFF module and SizeAdaptive loss can promote segmentation accuracy a lot compared with the baseline model. The results indicate that our proposed method can overcome the influence of morphological variety in some degree and is competitive with other current medical image segmentation algorithms. This novel methodology will be expected to use in clinic for image diagnosis in the future.

*Corresponding author.

ICMLC 2023, February 17–20, 2023, Zhuhai, China

ACM ISBN 978-1-4503-9841-1/23/02…$15.00
https://doi.org/10.1145/3587716.3587777

CCS CONCEPTS

• **Computing methodologies**; • **Computer vision problems**;

KEYWORDS

MRI, Lumbar Spine, Medical Image Segmentation, Size-Adaptive Loss

ACM Reference Format:

YINGDI, ZHANG, ZELIN, Shi, HUAN, Wang, CHONGNAN, Yan, LANBO, Wang, YUEMING, Mu, YUNPENG, Liu, SHUHANG, Wu, TIANCI, Liu, and MINGQI, Pang. 2023. Accurate Segmentation of Lumbar Spine Magnetic Resonance Imaging by Deep Neural Network with Size-Adaptive Loss. In *2023 15th International Conference on Machine Learning and Computing (ICMLC 2023), February 17–20, 2023, Zhuhai, China.* ACM, New York, NY, USA, 6 pages. https://doi.org/10.1145/3587716.3587777

1 INTRODUCTION

Lumbar spine anatomy segmentation plays an important role in radiological assessment and diagnosis in clinical settings[1]. Accurate and stable segmentation of the lumbar spine promotes understanding of the relationship between lumbar spine anatomy morphology or composition and lumbar diseases such as muscle degenerative changes, lumbar spinal stenosis and sarcopenia [1, 2]. Over recent decades, a variety of imaging modalities including magnetic resonance imaging (MRI), ultrasound and computed tomography (CT) have been used in the measurement of lumbar anatomy.

There is an considerable attention in the relationship between paraspinal muscle morphology or fatty infiltration and low back pain (LBP) [3]. MRI has been used to assess atrophy and fatty infiltration of muscles because it provides non-invasive and reproducible information on muscle cross-section areas. To date, several qualitative and quantitative approaches have been used to evaluate the grade of fatty infiltration. The visually grading based qualitative approaches show both good intra-observer and inter-observer agreement but poor precision therefor limiting clinical applicability. The muscle segmentation-based approaches can quantify the muscle size and fat content precisely while manual segmentation methods are tedious and rater-dependent which hamper further scientific and clinical applications [4]. Statistical shape modeling has been applied in segmenting quadratus lumborum with dice similarity of 0.87 which is relatively a low accuracy [4]. Automated thresholding algorithm and population-averaged atlas [5] have been proposed to be a basic technology of semi-automated or automated segmentation methods.

Expect for LPB, diagnosis of sarcopenia and lumbar central canal stenosis (LCCS) requires precise lumbar spinal anatomy segmentation [6]. The amount of muscle is a direct measurement of sarcopenia. Although both CT and MRI are considered to be very precise imaging systems, MRI provide better soft tissue contrast and is considered the gold standard for estimating muscle mass in research [4]. Therefore, accurate measurement of lumbar paraspinal muscle mass for diagnosing sarcopenia demands automated segmentation algorithm. LCCS is the most cause of disability in elderly and middle-aged patients in which the degenerative central canal narrows, due to degenerative disc bulging, hypertrophied ligamentum flavum, and degenerative changes in the facet joint. The separation degree of the cauda equina and morphology of the dural sac on T2-weighted axial images was the basis of grading LCCS [7], which had a high interobserver and intraobserver agreement. Dural sac cross-section area (DSCA) and spinal canal cross-section area (SCCSA) are another commonly used radiological measurement for LCCS. Recently, the dural sac area is founded a more sensitive parameter for evaluating LCCS than SCCSA. The computation of DSCA relies on precise automated dural sac segmentation algorithm.

The complex morphology of lumbar spine anatomy in T2-weighed images makes accurate segmentation a challenging task. Recently, deep learning based methods such as convolutional neural network (CNN) have enjoyed a great success in computer vision area [8]. With the widely prevalent of end-to-end fully convolutional network (FCN) [9], U-shape Net(U-Net) framework was proposed for medical image segmentation. Many U-Net based variations have proposed for different medical segmentation tasks such as vessel segmentation [10] and abdominal segmentation [11]. A U-shape model with attention mechanism was proposed for segmenting lumbar paraspinal muscles, while it cannot segment other anatomies of the lumbar spine [4].

A general limitation of the U-Net based models is that the reduced feature resolution caused by consecutive pooling operations or convolution striding. Although classification or object detection tasks take advantages from it, it often hinders segmentation tasks that require detailed spatial information. Maintaining high-resolution feature maps in the middle of the network is beneficial for segmentation tasks. However, high-resolution feature maps often impede accelerating the trainning process.

To avoid negative influence of different size anatomy, we propose a novel multi-scale feature fusion module (MSFF). The proposed MSFF could capture deeper and wider semantic features by infusing eight cascade branches with multi-scale atrous depthwise convolution or multi-scale image pooling operations. Besides, we propose a novel size-adaptive loss function varied from classical cross-entropy loss, which introduces an adaptive weighting factor for different class according to object size.

The main contributions of this work are summarized as follows:

1. To study automated lumbar spine anatomy segmentation method, we construct the first large-scale MRI lumbar spine segmentation dataset which contains 747 MRI images and corresponding anatomy masks. The expert segmentation masks with high labeling credibility, which can be regarded as ground truth, were labeled manually by several clinical experienced physicians.

2. We propose MSFF block to capture more high-level features and infuse multi-scale semantic features together, which boost segmentation performance of lumbar spine anatomy with different size and morphology.

3. To tackle the difficulties of segmenting lumbar spine anatomy with imbalanced size, we propose a novel loss function called size-adaptive loss which adaptively weights the loss to anatomy with different sizes.

2 MATERIALS AND METHODS

In this section, we present the framework for lumbar spine anatomy segmentation models. Figure.1 describe the proposed model in detail. The network takes the lumbar spine anatomy MRI image as input and outputs the segmentation result in an end-to-end manner. Similar to U-Net, the proposed architecture consists of four parts: the encoder, the proposed MSFF as bottleneck, the decoder and the concatenation operations.

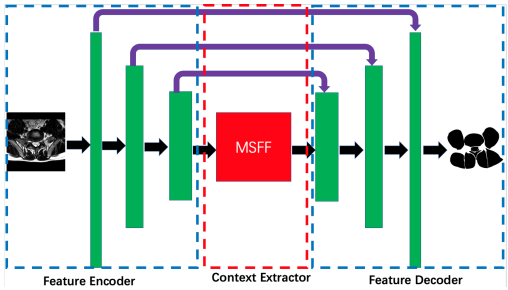

Figure 1: Illustration of the proposed network architecture. Firstly, the images are fed into a feature encoder module, where ResNet-18 pretrained from ImageNet is used to replace the original UNet encoder block. The context extractor is proposed to generate multi-scale feature maps. The decoder block contains 1×1 and 3×3 transposed convolution operations. Based on skip connections and the decoder block, we obtain the segmentation prediction map.

2.1 Feature Encoder Module

In U-Net architecture, each block of the encoder contains two convolution layers and one max pooling layer. In the proposed method, we replace it with the pretrained ResNet-18 in the feature encoder module [12], which reserve the first four feature extracting blocks without the last average pooling layer and the fully connected layers. Compared with the original U-Net encoder block, ResNet adds shortcut operations to prevent the gradient vanishing and is easier to optimize.

2.2 Multi-Scale Feature Fusion Module

The multi-scale feature fusion module is a novel proposed module, which consists of multi-scale atrous depthwise convolution branch and multi-kernel pooling branch. This module extracts multi-scale semantic information and generates more high level contextual abundant feature maps.

1)Atrous convolution: Deep convolution layers have shown to be successful in extracting feature representations in semantic segmentation and object detection tasks. However, atrous convolution [13], which is original proposed for the high-efficiency computation of wavelet transform and a generalization of the standard convolution operation, is a more powerful tool that allows us to explicitly control the resolution of features without increasing the number of parameters and adjust kernel's field-of-view by setting atrous rate in order to capture multi-scale information. In the case of two-dimensional signals, the atrous convolution is computed as follows:

$$y[i] = \sum_{k=1}^{K} x[i + r \times k] \times w[k] \tag{1}$$

where x[i] denotes the input signal, y[i] denotes the output signal, r denotes the atrous rate with which we sample the input signal, w[k] denotes the k-th parameter of the filter. Atrous convolution is equivalent to convolving of the input x by inserting $r - 1$ zeroes between two consecutive values of the convolution filter. Standard convolution is a special case for atrous rate r = 1, and we can

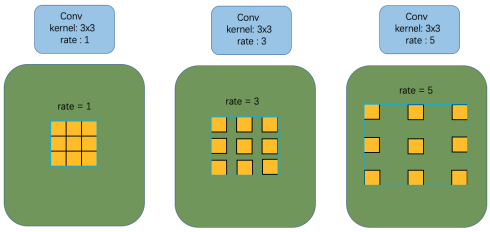

Figure 2: Atrous convolution with kernel size 3×3 and different dilation rates.

adaptively modify the filter's field-of-view by changing the atrous rate value. See Figure.2 for illustration.

2)Atrous depthwise separable convolution: Depthwise separable convolution, a powerful operation to reduce computation and parameters while having little performance loss, decompose a standard convolution operation into a depthwise convolution followed by a pointwise convolution. The size of the lumbar spine anatomy to be segmented in the MRI images is different. In order to segment the anatomy of different size, we replace standard convolution with depthwise separable convolution with different atrous rate which called atrous depthwise separable convolution(ADSConv).

The MSFF module consists of eight branches of three types. The first type of 3 branches are based on ADSConv. As shown in Fig.3 The ADSConvs and one 1×1 convolution layer are stacked in a parallel way. In this case, The three parallel branches with different atrous rate have different receptive fields. Intuitively, the convolution of large receptive field could extract more abstract features which has better performance for recognition of large object, while the convolution of small receptive field performs better in recognition of small objects. By combining different branches' output result, the MSFF module is able to extract features for segmenting objects with various size. In order to further reduce the computational complexity, we insert one 1×1 single channel convolution layer in front of every ADSConv.

A challenge in lumbar spine anatomy segmentation is the large variation of anatomy size in MRI image. In this paper, we propose a multi-kernel average pooling module to address this problem. The size of receptive field roughly determines the context information we can use. The general average pooling operation just apply a single pooling kernel like 2×2. In order to acquire multi-scale global context information, we employ four level global average pooling operation with kernel size of 2×2, 3×3, 4×4, 5×5 in a parallel way, as illustrated in Figure.3. The four-level outputs contains global feature with different receptive field and encodes different hierarchy global information. In order to reduce the dimension of weights and computational burden, we apply a 1×1 convolution after each average pooling operation. It reduces the dimension of the feature map. Next, we upsample the dimension-reduced features via bilinear interpolation to get the feature maps that have the same size as the original feature map. Finally, the multi-kernel average convolution output features are concatenated with ADSConv output features to get the MSFF bottleneck output feature map.

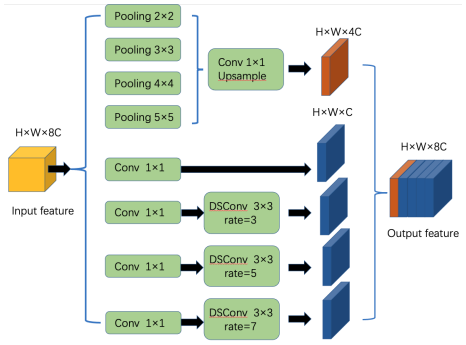

Figure 3: Multi-scale Feature Fusion Module.

2.3 Feature Decoder Module

The output feature of the proposed network is of the same size to input image. To restore the high level semantic features from the MSFF module, we thus propose a simple yet effective decoder module, as illustrated in Figure.1. Compared with linear interpolation operation, the transposed convolution operation could restore feature with more detailed information in a self-adaptive manner, which is more suitable in this case. So the repeated transposed convolution operations are adopted to restore the higher resolution feature, which is commonly used in the U-shape networks. Then every output of transposed convolution is concatenated with the corresponding encoder through the skip connection. The feature map from corresponding encoder contains more low level information which is crucial in accurate segmentation. Finally, the feature information is refined using two 3×3 convolution layers, and output a mask, which is the same size as the original input image.

2.4 Loss Function

Our framework is an end-to-end deep learning system. we need to train the proposed neural network to assign each pixel a class label. The most common loss function is cross-entropy loss function which is defined as:

$$L_{ce} = -\frac{1}{N}\sum_{i=1}^{i=N}\sum_{c=1}^{7} y_{o,c}\log\left(p_{o,c}\right) \tag{2}$$

where p represents predicted probability observation o is of class c. y is a binary indicator (0 or 1) if class label c is the correct classification for observation o. N indicates the total number of pixels in an image.

Anatomy size varies a lot in lumbar spine MRI images. This imbalance causes inefficient training because larger anatomical structure dominates the optimization process, which induces bad segmentation results of small anatomical structure. To address the problem of anatomy size imbalance, we propose a novel loss function called Size-Adaptive loss which is adapted from vallina cross-entropy loss. We introduce an adaptive weighting factor according to anatomy structure size for every class which amplifies the loss value of small anatomy and reduces the loss value of large anatomy adaptively. This reshaped loss function down-weight large size examples and up-weight small size examples thus focus training on hard small examples. The size-adaptive loss can be defined as follows:

$$L_{sa} = -\frac{1}{N}\sum_{i=1}^{i=N}\sum_{c=1}^{7}(1-S_c)^{\alpha} y_{o,c}\log(p_{o,c}) \tag{3}$$

Sc is the area of class c. α is focusing parameter.

Intuitively, the focusing parameter α smoothly adjusts the rate at which different anatomies are weighted. When $\alpha = 0$, Size-Adaptive loss is equivalent to cross-entropy loss. As α is increased the effect of the modulating factor is likewise increased. Experimental results suggest that $\alpha = 1$ work best.

2.5 Datasets

For validation of the proposed method, we construct the first lumbar spine anatomy segmentation dataset. The dataset contains 747 axial T2-weighted lumbar images of 83 male individuals, aged between 18-35 years old and obtained from Shengjing Hospital of China Medical University in Shenyang, China, which are randomly split into 603 for training and 144 for testing.

The Philips magnetic resonance was used, the repetition time of sagittal scanning is 2500ms, and that of axial scanning is 24855ms; the echo time of sagittal scanning is 80ms, that of axial scanning is 120ms and that of axial scanning is 4mm under 3.0T. The sagittal position nearest to the midline was selected as the location image. The axial images corresponding to L3-4, L4-5, L5-S1 discs were scanned. Each disc was divided into three slices. By excluding obvious disc herniation, infection, fracture, tumor and other abnormal changes and incomplete images, 747 T2-weighted axial images of 83 patients were obtained. All images are processed by brightness and contrast adjustment and normalized operation. To ensure the accuracy and validity of labeling, five spine surgeons and one imaging surgeon utilized Photoshop graphics software to label the bilateral erector spinae muscles(ES), bilateral multifidus muscles(MF), bilateral psoas major muscle(PS), disc, dural sac(DS), ligamentum flavum(LF) and lamina in the image manually.

A set of image and corresponding masks are illustrated in Fig.4. We can deduce from Figure.4 that the boundaries of lumbar spine anatomy are ambiguous especially the one between MF and ES.

To evaluate the segmentation results properly, we use dice similarity coefficient(DSC) as segmentation measurement, which shows similarity of region between results of the methods.

3 RESULTS

3.1 Implementation Details

All models are implemented in Ubuntu 16.04 system based on the deep learning library Pytorch. We use one NVIDIA 1080Ti GPU with 11 Gigabyte memory for both training and testing. In training phase, we adopt SGD algorithm to optimize the network with a batch size of 4. All models are trained for 100 epochs from scratch. The learning rate is initially set to 0.001 in first 80 epochs and decrease to 0.0001 in subsequent 20 epochs. We do data augmentation with random contrast, random brightness, random rotation in range [−10, 10] and random horizontal flip. When testing, we apply forward pass on the original image.

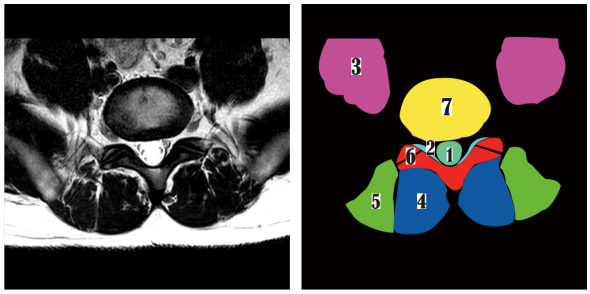

Figure 4: A set of image and corresponding masks in the proposed dataset.From 1 to 7, it is respectively dural sac, ligamentum flavum, bilateral psoas major muscles, bilateral multifidus muscles, bilateral erector spinae muscles, lamina and disc.

Table 1: Comparison of lumbar anatomy segmentation between different methods

Method	DSC↑
FCN	0.8258
U-Net	0.8339
SegNet	0.8384
AttentionUNet	0.8593
CE-Net	0.8633
LumNet	0.8727
Proposed	**0.8911**

3.2 Performance of the segmentation Task

In this subsection, we show the effectiveness of proposed method empirically. Some inference results are showed in Figure.5. The red region represents ground truth labeled by clinical experts, the green predicted results and the yellow region intersection of ground truth and inference results. As Figure.5 shows, the proposed method can have a desirable segmentation results even for obscure boundaries and small anatomy.

To demonstrate the advantage of our proposed method, we compare our method with several state-of-the-art algorithms including Fully Convolutional Network(FCN) [9], SegNet [14], U-Net [15], Attention U-Net [11], CE-Net [16] and LumNet [4]. We employ dice similarity coefficient (DSC) to quantitatively evaluate the segmentation performance.

Table 1 shows the metric results of different method. As we can see, the proposed method achieves 0.8911 DSC , which outperforms other state-of-the-art segmentation methods in a relatively large margin.

3.3 Experiment of Size-Adaptive loss

To analyze the influence of different α value in training procedure, we train the proposed models by size-adaptive loss with different α. As experiment in Table 2 shows, when α increase from 0 to 4 gradually the method performs better at the beginning and reaches the max DSC of 0.8911 in α is 1 and then gradually decays. We choose $\alpha = 1$ as the optimal value in other experiments. The proposed size-adaptive loss degenerates to cross-entropy loss in $\alpha = 0$. The experiment results indicate that the proposed size-adaptive loss boost training procedure of segmenting lumbar anatomy.

Table 2: Different α for Size-Adaptive loss

α	DSC↑
0	0.8871
0.1	0.8892
0.2	0.8888
0.5	0.8899
1	**0.8911**
2	0.8909
3	0.8888
4	0.8887

Table 3: Comparison of network with or without MSFF module

Method	DSC↑
network without MSFF	0.8706
network with MSFF	**0.8911**

3.4 Ablation study of using MSFF

As mentioned above, the proposed MSFF module which extracts multi-scale semantic information can boost segmentation performance. To evaluate the effect of MSFF, we conduct an experiment which make a comparison between network with or without MSFF. From the result of Table 3 we found that proposed MSFF module improve DSC by 2.05%, which demonstrate that the use of MSFF module lead to high-precision of lumbar anatomy segmentation.

4 DISCUSSION

Lumbar spine segmentation is important in medical image analysis and clinical settings. In this paper, we propose an end-to-end deep learning network for lumbar anatomy segmentation. Compared with U-Net, the proposed network adopts pretrained ResNet-18 in feature encoder block. A newly proposed Multi-Scale Fea- ture Fusion module as bottleneck is integrated into the network structure to capture more multi-scale semantic information. In order to solve the problem of multi-size anatomy segmentation, we recalibrate classic cross-entropy loss to size-adaptive loss. We are the first to realize automated segmentation of lumbar spine, which is a difficult problem because of various anatomy sizes and ambiguous anatomy boundary. Experiment results demonstrate that the using of MSFF and size-adaptive loss boost segmentation performance. In addition, the proposed method is a general solution and has the potential to be applied for other segmentation tasks.

5 CONCLUSION

The proposed algorithm can achieve accurate automatic segmentation of lumbar spine MRI, and has good clinical practicability. Our model can surmount the problem of anatomy variance in some

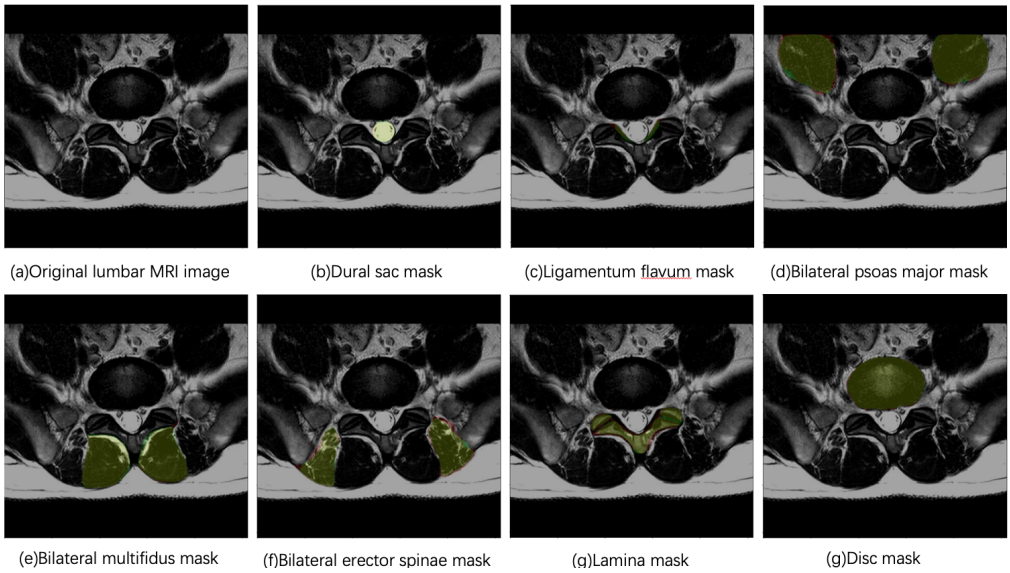

Figure 5: Lumbar spine anatomy segmentation results of proposed method. The red region denotes ground truth, and the green region denotes predicted anatomies from our method. The yellow region indicates pixels which are regarded as target anatomy by both ground truth and proposed method.

degree and competitive with other state of the art segmentation methods, which is crucial for lumbar spine disease diagnosis.

REFERENCES

[1] Lim, Y.S., Mun, J.-U., Seo, M.S., Sang, B.-H., Bang, Y.-S., Kang, K.N., Koh, J.W., Kim, Y.U.: Dural sac area is a more sensitive parameter for evaluating lumbar spinal stenosis than spinal canal area: A retrospective study. Medicine 96(49) (2017)

[2] Cruz-Jentoft, A.J., Baeyens, J.P., Bauer, J.M., Boirie, Y., Cederholm, T., Landi, F., Martin, F.C., Michel, J.-P., Rolland, Y., Schneider, S.M., *et al.*: Sarcopenia: European consensus on definition and diagnosisreport of the european working group on sarcopenia in older peoplea. j. cruz-gentoft *et al.* Age and ageing 39(4), 412–423 (2010) Martin A. Fischler and Robert C. Bolles. 1981. Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography. Commun. ACM 24, 6 (June 1981), 381–395. https://doi.org/10.1145/358669.358692

[3] Ghiasi, M.S., Arjmand, N., Shirazi-Adl, A., Farahmand, F., Hashemi, H., Bagheri, S., Valizadeh, M.: Cross-sectional area of human trunk paraspinal muscles before and after posterior lumbar surgery using magnetic resonance imaging. European Spine Journal 25(3), 774–782 (2016)

[4] Zhang, Y., Shi, Z., Wang, H., Yan, C., Wang, L., Mu, Y., Liu, Y., Wu, S., Liu, T.: Lumnet: A deep neural network for lumbar paraspinal muscles segmentation. In: Australasian Joint Conference on Artificial Intelligence, pp. 574–585 (2019). Springer

[5] Xiao, Y., Fortin, M., Batti é, M.C., Rivaz, H.: Population-averaged mri atlases for automated image processing and assessments of lumbar paraspinal muscles. European Spine Journal 27(10), 2442–2448 (2018)

[6] Ignasiak, D., Valenzuela, W., Reyes, M., Ferguson, S.J.: The effect of muscle ageing and sarcopenia on spinal segmental loads. European Spine Journal 27(10), 2650–2659 (2018)

[7] Schizas, C., Theumann, N., Burn, A., Tansey, R., Wardlaw, D., Smith, F.W., Kulik, G.: Qualitative grading of severity of lumbar spinal stenosis based on the morphology of the dural sac on magnetic resonance images. Spine 35(21), 1919–1924 (2010)

[8] Serte S, Serener A, Al-Turjman F. Deep learning in medical imaging: A brief review[J]. Transactions on Emerging Telecommunications Technologies, 2022, 33(10): e4080.

[9] Long, J., Shelhamer, E., Darrell, T.: Fully convolutional networks for semantic segmentation. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 3431–3440 (2015)

[10] Wang, W., Zhong, J., Wu, H., Wen, Z., Qin, J.: Rvseg-net: An efficient feature pyramid cascade network for retinal vessel segmentation. In: International Conference on Medical Image Computing and Computer-Assisted Intervention, pp. 796–805 (2020). Springer

[11] Schlemper, J., Oktay, O., Schaap, M., Heinrich, M., Kainz, B., Glocker, B., Rueckert, D.: Attention gated networks: Learning to leverage salient regions in medical images. Medical image analysis 53, 197–207 (2019)

[12] He, K., Zhang, X., Ren, S., Sun, J.: Deep residual learning for image recognition. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 770–778 (2016)

[13] Chen, L.-C., Papandreou, G., Kokkinos, I., Murphy, K., Yuille, A.L.: Deeplab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected crfs. IEEE transactions on pattern analysis and machine intelligence 40(4), 834–848 (2017)

[14] Badrinarayanan, V., Kendall, A., Cipolla, R.: Segnet: A deep convolutional encoder-decoder architecture for image segmentation. IEEE transactions on pattern analysis and machine intelligence 39(12), 2481–2495 (2017)

[15] Ronneberger, O., Fischer, P., Brox, T.: U-net: Convolutional networks for biomedical image segmentation. In: International Conference on Medical Image Computing and Computer-assisted Intervention, pp. 234–241 (2015). Springer

[16] Gu, Z., Cheng, J., Fu, H., Zhou, K., Hao, H., Zhao, Y., Zhang, T., Gao, S., Liu, J.: Ce-net: Context encoder network for 2d medical image segmentation. IEEE transactions on medical imaging 38(10), 2281–2292 (2019)