Mathematical Biosciences and Engineering

http://www.aimspress.com/journal/MBE

MBE, 20(9): 15999–16014.
DOI: 10.3934/mbe.2023713
Received: 18 May 2023
Revised: 09 July 2023
Accepted: 25 July 2023
Published: 04 August 2023

Research article

Spine MRI image segmentation method based on ASPP and U-Net network

Biao Cai[1], Qing Xu[1], Cheng Yang[2], Yi Lu[1], Cheng Ge[3], Zhichao Wang[1], Kai Liu[1], Xubin Qiu[2,*] and Shan Chang[1,*]

[1] Institute of Bioinformatics and Pharmaceutical Engineering, Jiangsu University of Technology, Changzhou 213001, China
[2] The Third Affiliated Hospital of Soochow University, Changzhou 213000, China
[3] Key Laboratory of Marine Drugs, Chinese Ministry of Education, School of Medicine and Pharmacy, Ocean University of China, Qingdao 266003, China

* **Correspondence:** Email: qiuxubinyiyi@sina.com, schang@jsut.edu.cn.

Abstract: The spine is one of the most important structures in the human body, serving to support the body, organs, protect nerves, etc. Medical image segmentation for the spine can help doctors in their clinical practice for rapid decision making, surgery planning, skeletal health diagnosis, etc. The current difficulty is mainly the poor segmentation accuracy of skeletal Magnetic Resonance Imaging (MRI) images. To address the problem, we propose a spine MRI image segmentation method, Atrous Spatial Pyramid Pooling (ASPP)-U-shaped network (UNet), which combines an ASPP structure with a U-Net network. This approach improved the network feature extraction by introducing an ASPP structure into the U-Net network down-sampling structure. The medical image segmentation models are trained and tested on publicly available datasets and obtained the Dice coefficient and Mean Intersection over Union coefficients with 0.866 and 0.755, respectively. The experimental results show that ASPP-UNet has higher accuracy for spine MRI image segmentation compared with other mainstream networks.

Keywords: ASPP; U-Net; spine; segmentation; DeepLabV3

1. Introduction

The spine is composed of 26 vertebrae, including 7 cervical, 12 thoracic, 5 lumbar, 1 sacrum and 1 coccyx. Under certain predisposing factors and the accelerated pace of life, specific groups of people are prone to spinal joint misalignment, disc herniation and osteophytes [1–3]. Computed Tomography

(CT) and Magnetic Resonance Imaging (MRI) are frequently used for the diagnosis and preoperative examination of various diseases in the clinic. For examinations of the spine, CT has the advantages of high sensitivity, fast imaging and low cost. Because CT has a certain amount of radiation, MRI can be used instead of CT when examining patients such as pregnant women. There are few studies on spinal MRI image segmentation compared with CT [4,5]. Medical image segmentation technology can be widely used in the medical field, such as focus segmentation: the focus's shape, size and location can be identified and quantified through the segmentation of medical images. Organ segmentation: Medical image segmentation can be used to segment organs such as livers, hearts and lungs. This is useful for medical procedures such as surgical planning and cancer treatment. Medical image segmentation technology can improve the efficiency of diagnosis and treatment. Segmentation of spinal images can improve the efficiency of diagnosis and treatment of spinal related diseases, and lay the foundation for remote diagnosis, surgical rehearsal, organ printing and cloud-based teaching.

With the development of medical image segmentation, the spine in MRI images can be segmented out, and doctors can more easily and quickly grasp the patient's spine. Early MRI image segmentation was often based on a priori knowledge and traditional segmentation techniques. The image segmentation results were derived using morphological techniques and watershed algorithms [6–8]. Morphological techniques and watershed algorithms are relatively basic segmentation algorithms, which are more suitable for relatively simple and low-featured image segmentation. The Watershed algorithm first transforms the image into a gradient map and uses the concept of contour lines to find a more suitable "height", which eventually completes the image segmentation operation. In some of the split tasks, this algorithm has the problems of low accuracy and poor generalization ability. Currently, some researchers have combined traditional algorithms and deep learning algorithms to solve medical image segmentation problems [8], However, the method still has some problems such as poor segmentation accuracy.

With the development of artificial intelligence, many machine learning and deep learning methods are used in spine image segmentation, such as U-Net [9] and Fully Convolutional Networks (FCN) [10]. Both these algorithms are proposed after CNN and take advantage of the weight sharing and low parametric number of convolutions to solve the problem of poor segmentation accuracy of traditional image segmentation tasks to some extent. In the field of medical image segmentation, the U-Net network is widely used and improved by many groups [11–15]. The U-Net network structure is symmetrical, and the segmentation effect is greatly improved compared with the traditional segmentation algorithm. In the subsequent image segmentation field, many networks are largely influenced by U-Net and FCN [16,17]. After U-Net was proposed, DeepLab series networks, nnU-Net networks [18] and SegNet [19] networks were proposed successively, and the segmentation accuracy was continuously improved. In the DeepLab series, the ASPP structure in DeepLabV3 [20] was proposed and first parallels different ratios of Dilated convolution and normal convolution (1 × 1). Then, it restores the data to the original size by splicing operation and normal convolution (1 × 1), and increasing the number of a few parameters while expanding the perceptual field.

In the field of medical image segmentation, the existing models may have some problems such as low segmentation accuracy or insufficient detail. We attempted to solve the above problem by adding an ASPP module to the U-Net network. Therefore, we proposed a medical image segmentation model named ASPP-UNet, which integrates multiple ASPP modules into the down-sampling process of the U-Net network. We believe that it is very important to improve the receptive field of medical image segmentation networks, and the ASPP module has a very significant advantage in this aspect. Compared with the traditional U-Net network, the ASPP-UNet model adds fewer parameters, but has a larger sensitivity field when extracting features, which can improve the accuracy of image segmentation. Our ASPP-UNet model not only increases the number of parameters, but can also significantly improve the accuracy of image segmentation, which is a great improvement compared with the traditional U-Net network. Compared with the DeepLabV3 network, our model has a larger drop in parameters with the same accuracy. Although the ASPP-UNet network is an obvious improvement, it also has some limitations in terms of accuracy improvement compared to the current mainstream segmentation networks. However, compared with the mainstream medical image segmentation network nnU-Net, our model also has some advantages. This study provides a new idea and solution for the algorithm design in the field of medical image segmentation.

2. Materials and methods

2.1. Dataset

The dataset was obtained from the second China Image and Graphics Society's (CIGS) Image and Graphics Technology Challenge, with 172 MRI data of the spine with annotation, Pang et al. used hybrid supervised learning to complete the task of spinal MRI image segmentation in this dataset, and achieved good segmentation results [21,22]. At present, the age, gender, disease and other information of the subjects corresponding to the 172 MRI image data is not known. The website is https://www.spinesegmentation-challenge.com. In this paper, the size of spinal MRI images using the publicly available data set is inconsistent, as shown in Table 1. Since most of the data size is 12 × 880 × 880, the size of all MRI images is unified to 12 × 880 × 880. As shown in Figure 1(a), spine MRI image segmentation tasks were performed on eight targets in this work, including sacral S, lumbar L1–L5 and thoracic T11 and T12, while thoracic T9 and T10 in some data was ignored.

Table 1. Original data size of spinal MRI images in data set.

Shapes	number
12, 880, 880	126
15, 880, 880	32
12, 1008, 1008	4
12, 512, 512	3
15, 960, 960	2
12, 864, 864	2
15, 896, 896	1
12, 960, 960	1
12, 1024, 1024	1

In order to reduce the amount of computation without degrading the accuracy, we removed the left and right full zero backgrounds of the data and labels. As shown in Figure 1(b) and (c), the data size is unified to 12 × 464 × 880. For this operation, we and Huang use the same dataset and data processing methods [23].

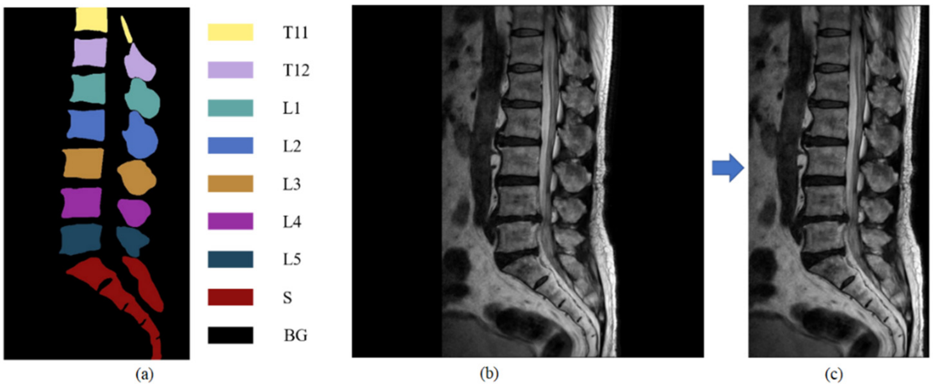

Figure 1. Schematic diagram of spine data processing and labeling. (a) is the label display after deleting part of the background, in which there are 9 labels, “BG” represents the background, “S” and “L” represent the sacrum and lumbar spine. (b) is the original data in the data set. (c) is the date with deleting some background.

2.2. ASPP module

In 2018, Chen et al. [20] used the atrous spatial pyramidal pooling module with holes (ASPP) and encoding-decoding structure for semantic segmentation. Low-level features for semantic segmentation of images are obtained by extracting network features from different classification networks. After normal convolution, ASPP is used to extend the width of the network and expand the field of perception by combining different ratios of dilated convolution and global pooling, etc. Furthermore, this operation will reduce the computation cost and allows the transforming of low-level features into high-level features. Using up-sampling such as bilinear interpolation, the size of high-level features is reduced to be the same as the size of low-level features. In the encoding and decoding process, the low-level features are combined with the up-sampled high-level features, and then the features are restored to the original image size using the up-sampling operation. During the segmentation process, the low-level features help the network to improve the segmentation details, and the high-level features help the network to improve the accuracy of the segmentation. In the subsequent semantic segmentation, many networks use the ASPP module [23,24].

Our work is inspired by the DeepLabV3 network, using the ASPP module shown in Figure 2 in the article. The speed of convergence and nonlinearity of the module is improved by using three dilated convolutions (the ratio of null convolutions is 6, 12 and 18), a 1 × 1 normal convolution and global average pooling in parallel. Batch Normalization (BN) and ReLU activation functions are added after the normal or dilated convolution. The parallel networks are stitched together by Concat. Normal convolution (convolution kernel is 1 × 1), BN and ReLU are used to ensure that the output image size is the same as the input image size.

ASPP is stitched together by five parallel lines, among which three lines use atrous convolution with different ratios. Atrous convolution and ordinary convolution have the same number of parameters, but they have obvious advantages in the receptive field. This is the fundamental reason why the ASPP module has great advantages in expanding the receptive field, and why we choose ASPP and U-Net networks to combine. The experiment proves that our choice is reasonable in the direction of spinal MRI image segmentation.

The DeepLabV3 network is typically divided into three steps using a different encoder to extract features. The ASPP module increases the sensitivity field, and then linear interpolation completes the up-sampling. In contrast, the ASPP-UNet network is a little different. First, during the down-sampling process, we used the ASPP module several times to make the network have a larger receptive field. In addition, ASPP-UNet used the same jump join structure as U-Net between up-sampling and down-sampling, allowing more segmentation details to be preserved.

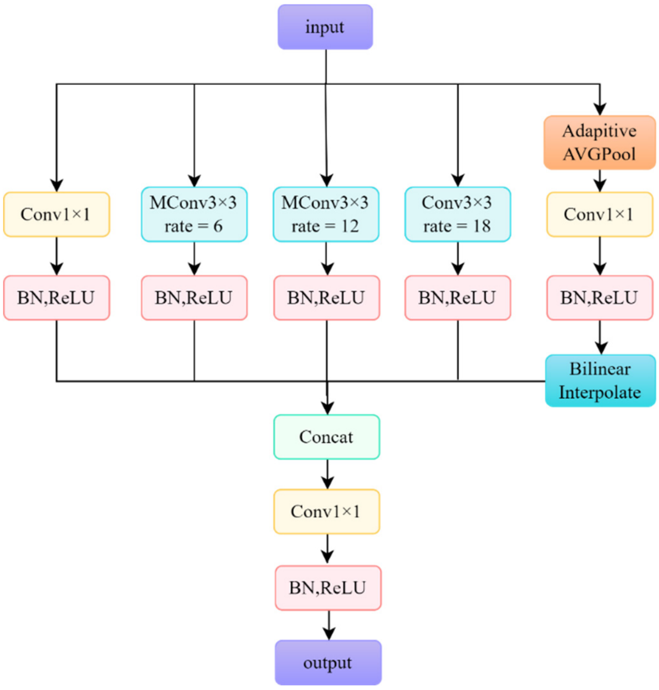

Figure 2. Schematic diagram of ASPP network. “BN” represents Batch Normalization; “Conv” represents normal convolution, and the subsequent numbers represent the size of the convolution kernel. “MConv” represents the dilated convolution, the subsequent number represents the size of the dilated convolution kernel and “rate” represents the ratio of dilated convolution. “Adaptive AVGPool” represents the global average pooling. “Bilinear Interpolate” represents bilinear interpolation.

2.3. DC module

In order to better extract high-level features for medical image segmentation, the ASPP module is added to the U-Net network down-sampling process for feature extraction. As shown in Figure 3, we designed the Double Convolution (DC) module and used it for the down-sampling process of the network. The major structure of the module is the normal convolution (convolution kernel size is 3 × 3), BN and ReLU and repeated twice, which is used to improve the extraction of high-level features from the network while expanding the network perceptual field.

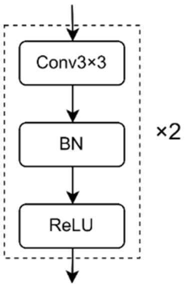

Figure 3. DC module. “BN” represents Batch Normalization; “Conv” represents normal convolution, and the subsequent numbers represents the size of the convolution kernel. “ReLU” represents the ReLU Activation function.

2.4. Network architecture example

In 2015, the U-Net network [9] was proposed and widely used in the field of image segmentation. Although there are many variations and optimizations for the U-Net network, this network structure is still one of the classical image segmentation networks [25–28]. U-Net is named after the shape of its network, which is similar to the letter “U”. In the network structure, the structure is symmetrically divided into an up-sampling part and a down-sampling part. Three modules are combined into the down-sampling part, including two convolution operations, one pooling operation and ReLU. The feature size is continuously reduced and high-level features are obtained using multiple down-sampling parts. In the up-sampling part, the high-level features are reduced to the original image size using deconvolution and linear interpolation. The addition of several skip connections between down-sampling and up-sampling in order to enhance the details of network segmentation and speed up the convergence of the network. The U-Net network has many advantages, such as a relatively simple network structure, easy training with fewer parameters and easy control of the network size.

The specific architecture of the network is shown in Figure 4. In the network structure, it can be divided into two parts, the left side is the down-sampling part and the right side is the up-sampling part. In the down-sampling process, the DC module is first passed once, and then the DC module, ASPP module and MaxPool are operated four times in turn. In addition, after each DC module, the obtained feature data is transmitted to the up-sampling section to complete the stitching operation. In the up-sampling part, after a transpose convolution operation, and then going through Concat, DC and TranConv operations three times in turn. Each time, Concat completes the splicing operation with the data transmitted by down-sampling. Finally, after going through the DC module once, the network structure of the whole module is completed.

2.5. Loss function

The training objective of the network contains a loss function, and the cross-entropy loss function is defined below.

$$Loss = -\frac{1}{un}\sum_{i=1}^{un} y_{true}\log\left(y_{pre}\right) \tag{1}$$

Here, y_{true} represents the value of the pixel in the real label, and y_{pre} represents the value of the predicted result pixel, n represents the number of categories of labels (n = 9) and u represents the number of pixel points in an image.

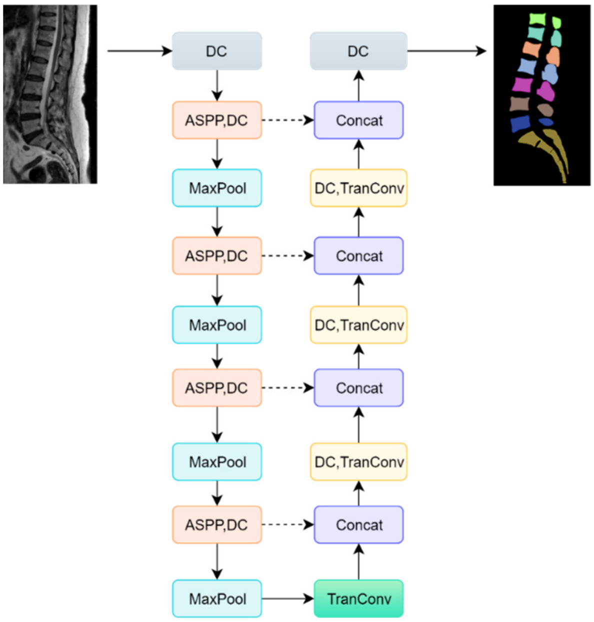

Figure 4. Overall network architecture. “DC” represents the DC module, and “MaxPool” represents maximum pooling. “Concat” represents splicing operation, “TranConv” represents transposition convolution operation. “ASPP” represents the ASPP module, and the dotted line represents the data transfer operation.

2.6. Evaluation indicators

The main evaluation metrics used in this paper are Dice coefficient (DSC) and Mean Intersection over Union (MIoU), and the definitions of two rating metrics are as follows.

$$DSC_i = \frac{1}{n}\sum_{i=1}^{n} 2\frac{V_{gt_i}\cap V_{pre_i}}{V_{gt_i}+V_{pre_i}} \tag{2}$$

$$MIoU = \frac{1}{n}\sum_{i=1}^{n}\frac{V_{gt_i}\cap V_{pre_i}}{V_{gt_i}\cup V_{pre_i}} \tag{3}$$

where n represents the number of label types for image segmentation, and V_{gt_i} is the set of pixel points contained in the i-th real label, and V_{pre_i} is the set of pixel points predicted by the i-th model.

2.7. Details about some parameters

In this paper, the pytorch deep learning framework is used for training on a Tesla K80 machine with 11G video memory. The model optimization function is the stochastic gradient descent optimizer (SGD). The activation function is ReLU, the Batch Size is 2 and the training batch is 50. The learning rate is 0.01, and the learning rate variation strategy formula is shown below.

$$Ir_i = Ir_{i-1} \times 0.99 \tag{4}$$

where Ir_i is the learning rate of the ith training batch, and Ir_{i-1} is the learning rate of the *i-1st* training batch.

2.8. Network settings

We used a total of 6 medical image segmentation methods or network models for comparison. The specific configuration of segmentation methods is as follows:

- U-Net-2D: After converting both 3D MRI images and labels into multiple 2D images, the U-Net network is used to train images and labels and its scale is consistent with ASPP-UNet.
- DeepLabV3:The MRI dataset above is trained using the classical DeepLabV3 network, where the network size and parameters were not changed. Resnet101 network was selected as the feature extraction network for down-sampling.
- ASPP-UNet: The specific structure and parameters of the network have been described above.
- U-Net-3D: In order to construct 3D U-Net networks, the convolution and pooling of 2D U-Net networks are replaced by 3D convolution and 3D pooling. 2D U-Net networks and 3D U-Net networks have the same size and hyperparameters.
- nnU-Net-2D: It is a medical image segmentation method based on the U-Net network, which can automatically perform data pre-processing, training and post-processing. In addition, the network is trained with 1000 epochs.
- SegNet: The structure of the SegNet network we use has not changed. The size of the network is equivalent to that of ASPP-UNet and some of its super parameters are also consistent, such as batch size, Learning rate, etc.

To better evaluate the performance of the six network models, the network models are cross-validated fivefold on the training set data. Each segmentation method yields five models and segmentation results. We evaluated the network model using DSC and MIoU evaluation indicators, and all results were derived from the average of five cross-validations in order to reduce error.

2.9. Ablation experiment

Although the current ASPP-UNet network is designed, we conducted ablation experiments at the beginning of completion, mainly focusing on the location and number of ASPP modules. With other settings exactly the same, changing only the number and position of ASPP modules, we get DSC and MIoU for different ablation experimental models.

In Figure 4, the ASPP-UNet network has four ASPP modules, which we name ASPP1, ASPP2, ASPP3 and ASPP4. Ablation experiments designed for the presence or absence of four ASPPs are shown in Table 2. We first designed four sets of experiments to observe the DSC and MIoU accuracy of spinal MRI image segmentation present in only one ASPP module at four locations. Through the results, it can be found that the segmentation accuracy of ASPP increases somewhat with the deepening of position, and in the first four experiments, only ASPP4 has the highest segmentation accuracy. Moreover, we designed experiments using multiple ASPP modules (experiments 5, 6 and 7 in Table 2), and finally experiment 7 used the segmentation results of four ASPP modules with the best segmentation accuracy among all experiments.

Table 2. DSC and MIoU for ablation experiments for ASPP-UNet networks.

Index	ASPP1	ASPP2	ASPP3	ASPP4	MIoU (%)	DSC (%)
1	√	×	×	×	68.12 ± 1.12	80.18 ± 3.15
2	×	√	×	×	69.51 ± 2.57	81.31 ± 2.18
3	×	×	√	×	70.41 ± 2.16	81.92 ± 2.16
4	×	×	×	√	71.31 ± 2.65	82.60 ± 1.79
5	×	×	√	√	72.34 ± 2.28	83.90 ± 2.79
6	×	√	√	√	74.34 ± 2.26	82.94 ± 2.15
7	√	√	√	√	75.49 ± 1.98	86.60 ± 2.19

Finally, the ASPP-UNet network structure used in this paper is the structure of Experiment 7 in Table 2. The results show that using more ASPP modules or using ASPP in the deep network structure may improve the segmentation accuracy.

2.10. Generalization testing

To test the generalization ability of the model, we add noise to the spine MRI image data in the test set and use the trained model to infer the data with added noise. The calculation formula for adding noise is as follows.

$$Gray_{out} = Gray_{in} + k * Z \tag{5}$$

where $Gray_{in}$ and $Gray_{out}$ are the grayscale values of each pixel point in the spine MRI image before and after adding noise, respectively; k is the controllable coefficient of the degree of adding noise; Z is a random number that conforms to the standard normal distribution, and the Z corresponding to each pixel point needs to be regenerated randomly. Also, to ensure that $Gray_{out}$ is an integer between 0 and 255, we restrict $Gray_{out}$ accordingly to meet the requirements as an image.

After processing by adding noise, the boundary between the background and the label in the spine MRI image will be gradually blurred, which requires higher segmentation generalization ability. In Figure 5, we show the preprocessed MRI image, labels and prediction results of an image in the test set, and will add noise to the MRI image and make predictions. We add noise processing to one image in the test set with the values of 10, 30, 50, 70 and 90 in Eq 5 to generate five images with added noise, and after inference by the trained model, the results shown in Figure 6 are obtained. From Figure 6, it can be concluded that the inference ability of the model decreases continuously with the increase, and the generalization ability of the model is relatively limited. When it is greater than 30, the inference ability of the model decreases rapidly.

It can be found that the generalization ability of the model is relatively average. After discussion, we think there may be several reasons. First, the model was trained without adding noise, and the existence of a certain consistency of the data limited the generalization ability of the model. Second, the deep learning models of medical images generally have a certain problem of poor generalization ability. After that, we will start a study on noisy training of spine MRI images, and adding noise to the images of the training set may be helpful for the generalization ability.

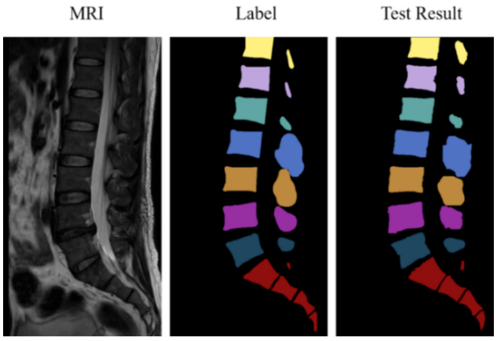

Figure 5. MRI images, labels and prediction results that increase the generalization ability of noise test models.

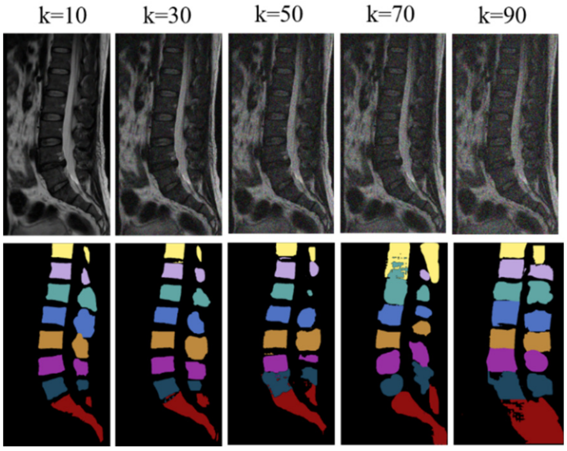

Figure 6. Increase MRI images and predictions with varying degrees of noise (The first row is an MRI image of the spine, the upper k is the controllable coefficient for increasing noise in Equation 5, and the second row corresponds to the prediction result).

3. Results

In order to compare our model with other methods, among 172 MRI images, we randomly divided the data set into a training set and a test set in a ratio of 4:1, with 138 pieces of data in the training set and 34 pieces of data in the test set. The training and testing of the model were completed 5 times cross-validation. In the training process of each fold, there is no overlap between the training set and the test set.

We compare the ASPP-UNet method with other classical methods, including U-Net-2D, DeepLabV3, SegNet, U-Net-3D and nnU-Net-2D. As shown in Figure 7, by predicting the same three subjects' MRI images, our algorithm is compared with the other five mainstream algorithms in terms of the details of image segmentation. In the collection of tags "BG", there is a small chance that our method incorrectly splits "BG" into other tags. For example, in the first row of Figure 7, some "BG" pixels in the U-Net-3D and nnU-Net methods are incorrectly divided into "L3" and "L4". In the set of labels "L1–L5", "T11–12" and "S", our method is unlikely to misclassify one label as another. For example, in the second row of Figure 7, the "L2" pixels of U-Net-3D and U-Net-2D methods are incorrectly divided into “L1” pixels, and a small number of “L3” pixels in U-Net-2D are divided into “L2” pixels. In the third row of Figure 5, all the methods split better, but ours has fewer “BG” pixels divided into labels. The SegNet model has more “BG” split into “L3” and L4. At the same time, the overall segmentation effect of the SegNet model is poor.

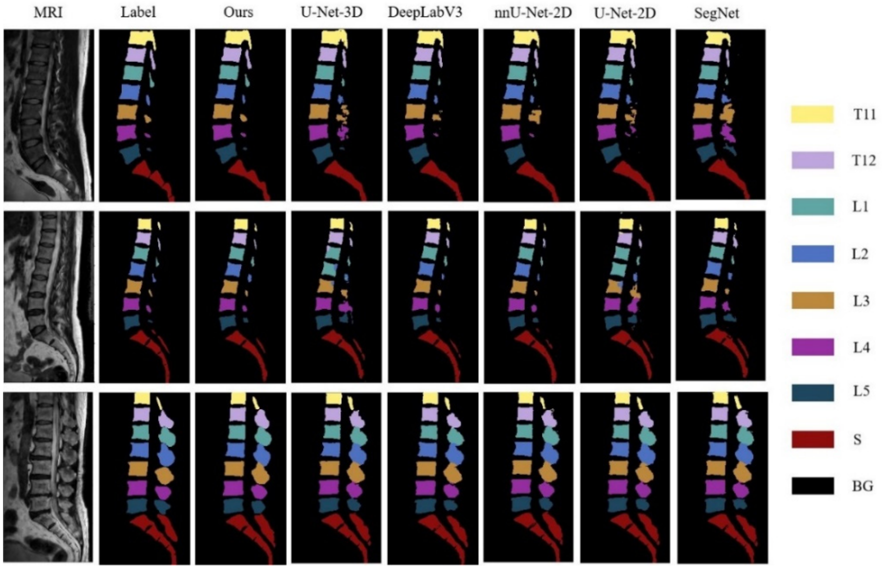

Figure 7. Visualization results between different methods, where each row represents one subject MR data and “BG” represents the background.

As shown in Figure 7, the reason for errors in other segmentation networks may be that the receptive field of the network is too small. Although the receptive field can be increased by increasing the depth of the network, the amount of computation will be greatly increased at the same time.

Table 3. Average DSC (%) and MIoU (%) of the vertebrae by each method.

	Model	S	L5	L4	L3	L2	L1	T12	T11	Mean
	U-Net-2D	85.0 ± 1.3	88.5 ± 2.4	87.4 ± 2.3	78.5 ± 4.2	67.5 ± 2.1	75.8 ± 3.2	78.4 ± 1.5	78.4 ± 4.2	79.94 ± 2.65
	SegNet	84.2 ± 1.9	87.0 ± 1.9	83.2 ± 2.2	81.6 ± 2.3	70.5 ± 1.9	77.1 ± 1.2	77.9 ± 1.5	78.3 ± 3.1	79.98 ± 2.01
	DeepLabV3	87.5 ± 2.1	86.5 ± 2.2	85.0 ± 2.5	80.5 ± 3.2	75.3 ± 3.5	83.4 ± 1.5	81.5 ± 1.2	84.0 ± 3.1	82.96 ± 2.41
	U-Net-3D	86.3 ± 1.5	89.5 ± 2.1	88.5 ± 3.2	82.8 ± 2.8	77.5 ± 2.1	85.7 ± 2.6	83.4 ± 2.5	84.8 ± 2.1	84.81 ± 2.36
	nnU-Net-2D	85.2 ± 1.2	85.5 ± 3.2	86.7 ± 2.4	86.4 ± 2.1	84.4 ± 1.9	81.8 ± 1.5	83.7 ± 1.6	89.0 ± 2.4	85.34 ± 2.04
DSC	ours	88.4 ± 1.5	91.0 ± 2.5	89.1 ± 2.4	87.5 ± 2.5	84.0 ± 2.3	82.0 ± 2.1	82.7 ± 1.4	88.1 ± 2.8	86.60 ± 2.19
(%)	U-Net-2D	73.5 ± 0.9	78.3 ± 1.5	76.3 ± 2.9	64.5 ± 1.6	54.2 ± 2.5	60.8 ± 2.8	66.2 ± 3.4	64.9 ± 2.5	67.34 ± 2.26
	SegNet	74.7 ± 1.5	78.5 ± 1.7	75.6 ± 1.4	66.7 ± 1.9	52.0 ± 2.9	61.7 ± 1.4	66.0 ± 3.9	62.9 ± 1.8	67.26 ± 2.06
	DeepLabV3	79.2 ± 0.5	76.1 ± 2.3	74.1 ± 2.3	68.5 ± 2.1	64.2 ± 3.4	71.4 ± 2.6	68.1 ± 2.9	73.1 ± 1.8	71.84 ± 2.24
	U-Net-3D	78.2 ± 0.8	81.5 ± 2.1	81.4 ± 2.5	71.4 ± 1.2	64.3 ± 2.1	74.8 ± 2.7	73.4 ± 2.7	74.4 ± 2.5	74.93 ± 2.08
	nnU-Net-2D	78.2 ± 1.2	79.4 ± 1.7	75.9 ± 1.5	77.5 ± 0.7	72.3 ± 4.5	68.9 ± 1.8	68.4 ± 2.4	78.7 ± 1.6	74.83 ± 1.93
	ours	79.1 ± 0.9	80.8 ± 1.2	78.0 ± 1.9	78.3 ± 1.4	72.4 ± 3.4	68.3 ± 2.6	69.1 ± 3.0	77.9 ± 1.4	75.49 ± 1.98

In addition, we also obtained the DSC and MIoU scores of each method, as shown in Table 3. The DSC and MIoU of the ASPP-UNet network were respectively 86.6% and 75.5%. Compared with U-Net-2D, SegNet, DeepLabV3, U-Net-3D and nnU-Net-2D, ASPP-UNet leads in DSC evaluation metrics by 6.66, 6.62, 3.64, 1.79 and 1.26%, respectively, and MIoU evaluation metrics by 8.15, 8.23, 3.65, 0.56 and 0.66%, respectively. On the eight spinal skeletal labels, the ASPP-UNet network leads by four DSC evaluation metrics, S, L5, L4 and L3, and three MIoU evaluation metrics, S, L3 and L2. The numbers after "±" in Table 3 represent standard deviations.

We compared our model with other models and calculated the number of parameters and inference time of the model. U-Net-2D, SegNet, U-Net-3D and our model use the same network scale. The model network depth is 4, and the number of channels is 64, 128, 256 and 512. DeepLabV3 uses ResNet101 as the feature extraction network, while nnU-Net uses a classical model setup. As shown in Table 4, the number of parameters and inference time in the ASPP-Unet network is relatively small, only slightly higher than that in the U-Net-2D network.

Table 4. Parameters and inference schedules for different networks.

Model	Params	Inference time(s/image)
U-Net-2D	15.08M	0.163s
SegNet	15.21M	0.170s
DeepLabV3	46.9M	0.198s
U-Net-3D	31.03M	0.221s
nnU-Net-2D	60.5M	0.292s
ours	22.06M	0.171s

In order to better illustrate the stability of the model, the Loss curve of the ASPP-UNet network and other networks in the training and testing process is given, as shown in Figures 8–9. We trained all models with 50 epochs, among which the nnU-Net network trained with 1000 epochs, and sampled Loss every 20 epochs. In Figure 8, we can see that the ASPP-UNet loss converges well and the curve flattens out after epoch 15. The curve is basically stable at epoch 35. When compared with other models, the ASPP-UNet network converges faster. In Figure 9, the difference between all models is not very big. In this work, there are only 42 cases of images in the test set, and there is slight Loss fluctuation in the test process. More spinal MRI data will be added to training and testing models in the future.

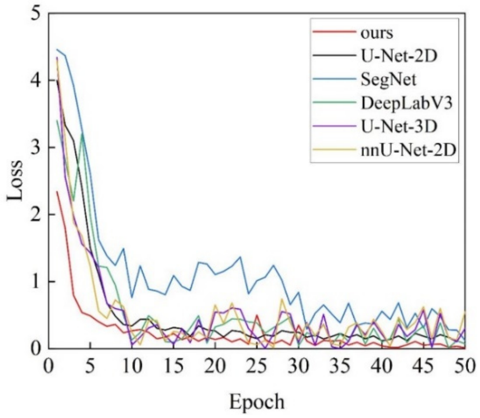

Figure 8. Loss curves of each model during training.

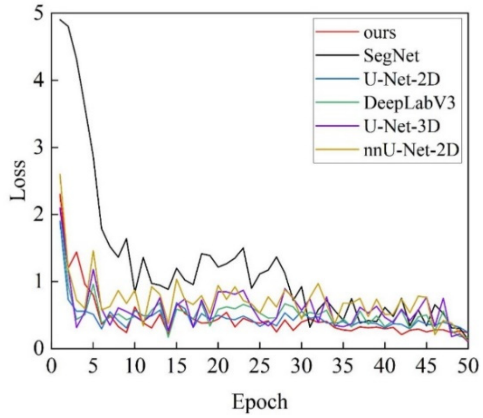

Figure 9. Loss curves of each model during testing.

4. Discussion

In this paper, we propose a segmentation method called ASPP-UNet for spine MRI images. This method combines U-Net and ASPP modules in DeepLabV3 and adds several ASPP modules to the part of the U-Net network down-sampled. It expands the model perceptual field without substantially increasing the number of model parameters. In addition, the structure of the up-sampling part of ASPP-UNet is exactly the same as that of U-Net.

Our main innovation in this work lies in the fusion of ASPP and U-Net networks, which has led to improved segmentation accuracy compared to the traditional U-Net network. Importantly, this improvement was achieved without significantly increasing the number of network parameters. Additionally, our approach exhibits outstanding performance in capturing segmentation details compared to some mainstream networks.

In future work, we will attach importance to the innovative nature of the article. By building upon the current research, we strive to make significant contributions and push the boundaries of innovation in this area.

Limitations of this study include the small size of the dataset used in our experiments. Although we carefully curated the dataset to ensure its relevance and diversity, a larger dataset would provide a more comprehensive evaluation of our proposed approach. Furthermore, we acknowledge that we did not compare our approach with a wider range of existing methods, which could have provided a more thorough analysis of its strengths and weaknesses.

Future work could focus on expanding the dataset used in this study to include more diverse examples and improving the evaluation by comparing our approach with a wider range of state-of-the-art methods. Additionally, further investigations into the underlying mechanisms of our proposed approach could help to provide a more nuanced understanding of its performance and potential limitations. With the continuous progress of technology and the accumulation of data, we believe that the performance and efficiency of medical image segmentation will be greatly improved.

Use of AI tools declaration

The authors declare they have not used Artificial Intelligence (AI) tools in the creation of this article.

Acknowledgments

This study was supported by NO.2021-NCRC-CXJJ-ZH-04 of Clinical Application-oriented Medical Innovation Foundation from National Clinical Research Center for Orthopedics, Sports Medicine & Rehabilitation and Jiangsu China-Israel Industrial Technical Research Institute Foundation.

The data set for this study was supported by the Chinese Society of Image and Graphics.

Conflict of interest

The authors declare there is no conflict of interest.

Ethics approval and consent to participate

All methods of this study were carried out in accordance with relevant guidelines and regulations. All experimental schemes were licensed and approved by the Graphics Society of China, and all participants gave informed consent to all contents of this paper.

References

1. D. Lee, S. H. Tak, Fear of falling and related factors in older adults with spinal diseases, *J. Gerontol. Nurs.*, **47** (2021), 29–35. https://doi.org/10.3928/00989134-20210624-05
2. F. C. Kohler, P. Schenk, M. Bechstedt-Schimske, B. W. Ullrich, F. Klauke, G. O. Hofmann, et. al., Open versus minimally invasive fixation of thoracic and lumbar spine fractures in patients with ankylosing spinal diseases, *Eur. J. Trauma Emerg. Surg.*, **48** (2021), 2297–2307. https://doi.org/10.1007/s00068-021-01756-3
3. F. R. V. Tol, A. L. Versteeg, H. M. Verkooijen, F. C. Öner, J. J. Verlaan, Time to surgical treatment for metastatic spinal disease: Identification of delay intervals, *Global Spine J.*, **13** (2021). https://doi.org/10.1177/2192568221994787
4. W. Jung, S. S. Shim, K. Kim, CT findings of acute radiation-induced pneumonitis in breast cancer, *Br. J. Radiol.*, **94** (2021). https://doi.org/10.1259/bjr.20200997
5. S. Amiri, M. Akbarabadi, F. Abdolali, A. Nikoofar, A. J. Esfahani, S. Cheraghi, Radiomics analysis on CT images for prediction of radiation-induced kidney damage by machine learning models, *Comput. Biol. Med.*, **133** (2021), 104409. https://doi.org/10.1016/j.compbiomed.2021.104409
6. H. Zhang, Z. Tang, Y. Xie, X. Gao, Q. Chen, A watershed segmentation algorithm based on an optimal marker for bubble size measurement, *Measurement*, **138** (2019), 182–193. https://doi.org/10.1016/j.measurement.2019.02.005
7. A. Kornilov, I. Safonov, I. Yakimchuk, A review of watershed implementations for segmentation of volumetric images, *J. Imaging*, **8** (2022), 127. https://doi.org/10.3390/jimaging8050127
8. A. Kucharski, A. Fabijańska, CNN-watershed: A watershed transform with predicted markers for corneal endothelium image segmentation, *Biomed. Signal Process.*, **68** (2021), 102805. https://doi.org/10.1016/j.bspc.2021.102805

9. O. Ronneberger, P. Fischer, T. Brox, U-net: Convolutional networks for biomedical image segmentation, in *International Conference on Medical Image Computing and Computer-Assisted Intervention*, **9351** (2015), 234–241.
10. J. Long, E. Shelhamer, T. Darrell, Fully convolutional networks for semantic segmentation, in *2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, (2015), 3431–3440. https://doi.org/10.1109/CVPR.2015.7298965
11. J. Zhang, C. Li, S. Kosov, M. Grzegorzek, K. Shirahama, T. Jiang, et. al., LCU-Net: A novel low-cost U-Net for environmental microorganism image segmentation, *Pattern Recognit.*, **115** (2021), 107885. https://doi.org/10.1016/j.patcog.2021.107885
12. Z. Liu, Y. Cao, Y. Wang, W. Wang, Computer vision-based concrete crack detection using U-net fully convolutional networks, *Autom. Constr.*, **104** (2019), 129–139. https://doi.org/10.1016/j.autcon.2019.04.005
13. J. Zhou, Y. Lu, S. Tao, X. Cheng, C. Huang, E-Res U-Net: An improved U-Net model for segmentation of muscle images, *Expert Syst. Appl.*, **185** (2021), 115625. https://doi.org/10.1016/j.eswa.2021.115625
14. X. Dong, Y. Lei, T. Wang, M. Thomas, L. Tang, W. J. Curran, et. al., Automatic multiorgan segmentation in thorax CT images using U-net-GAN, *Med. Phys.*, **46** (2019), 2157–2168. https://doi.org/10.1002/mp.13458
15. G. Tong, Y. Li, H. Chen, Q. Zhang, H. Jiang, Improved U-NET network for pulmonary nodules segmentation, *Optik*, **174** (2018), 460–469. https://doi.org/10.1016/j.ijleo.2018.08.086
16. N. Siddique, S. Paheding, C. P. Elkin, V. Devabhaktuni, U-net and its variants for medical image segmentation: A review of theory and applications, *IEEE Access*, **9** (2021), 82031–82057. https://doi.org/10.1109/ACCESS.2021.3086020
17. G. Du, X. Cao, J. Liang, X. Chen, Y. Zhan, Medical image segmentation based on u-net: A review, *J. Imaging Sci. Technol.*, **64** (2020), 1–12. https://doi.org/10.2352/J.ImagingSci.Technol.2020.64.2.020508
18. H. El-Hariri, L. A. S. M. Neto, P. Cimflova, F. Bala, R. Golan, A. Sojoudi, et. al., Evaluating nnU-Net for early ischemic change segmentation on non-contrast computed tomography in patients with Acute Ischemic Stroke, *Comput. Biol. Med.*, **141** (2022), 105033. https://doi.org/10.1016/j.compbiomed.2021.105033
19. V. Badrinarayanan, A. Kendall, R. Cipolla, Segnet: A deep convolutional encoder-decoder architecture for image segmentation, *IEEE Trans. Pattern Anal. Mach. Intell.*, **39** (2017), 2481–2495. https://doi.org/10.1109/TPAMI.2016.2644615
20. L. Chen, G. Papandreou, I. Kokkinos, K. Murphy, A. L. Yuille, Deeplab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected crfs, *IEEE Trans. Pattern Anal. Mach. Intell.*, **40** (2017), 834–848. https://doi.org/10.1109/TPAMI.2017.2699184
21. S. Pang, C. Pang, L. Zhao, Y. Chen, Z. Su, Y. Zhou, et. al., SpineParseNet: Spine parsing for volumetric MR image by a two-stage segmentation framework with semantic image representation, *IEEE Trans. Med. Imaging*, **40** (2020), 262–273. https://doi.org/10.1109/TMI.2020.3025087
22. S. Pang, C. Pang, Z. Su, L. Lin, L. Zhao, Y. Chen, et. al., DGMSNet: Spine segmentation for MR image by a detection-guided mixed-supervised segmentation network, *Med. Image Anal.*, **75** (2022), 102261. https://doi.org/10.1016/j.media.2021.102261

23. R. Liu, F. Tao, X. Liu, J. Na, H. Leng, J. Wu, et. al., RAANet: A residual ASPP with attention framework for semantic segmentation of high-resolution remote sensing images, *Remote Sens.*, **14** (2022), 3109. https://doi.org/10.3390/rs14133109
24. T. Lei, R. Wang, Y. Zhang, Y. Wan, C. Liu, A. K. Nandi, DefED-Net: Deformable encoder-decoder network for liver and liver tumor segmentation, *IEEE Trans. Radiat. Plasma Med. Sci.*, **6** (2021), 68–78. https://doi.org/10.1109/TRPMS.2021.3059780
25. Y. Weng, T. Zhou, Y. Li, X. Qiu, Nas-unet: Neural architecture search for medical image segmentation, *IEEE Access*, **7** (2019), 44247–44257. https://doi.org/10.1109/ACCESS.2019.2908991
26. Z. Luo, Y. Zhang, L. Zhou, B. Zhang, J. Luo, H. Wu, Micro-vessel image segmentation based on the AD-UNet model, *IEEE Access*, **7** (2019), 143402–143411. https://doi.org/10.1109/ACCESS.2019.2945556
27. P. Ahmad, H. Jin, R. Alroobaea, S. Qamar, R. Zheng, F. Alnajjar, et. al., MH UNet: A multi-scale hierarchical based architecture for medical image segmentation, *IEEE Access*, **9** (2021), 148384–148408. https://doi.org/10.1109/ACCESS.2021.3122543
28. X. Li, H. Chen, X. Qi, Q. Dou, C. Fu, P. Heng, H-DenseUNet: hybrid densely connected UNet for liver and tumor segmentation from CT volumes, *IEEE Trans. Med. Imaging*, **37** (2018), 2663–2674. https://doi.org/10.1109/TMI.2018.2845918