International Journal of Machine Learning and Cybernetics (2023) 14:3671–3683
https://doi.org/10.1007/s13042-023-01857-y

ORIGINAL ARTICLE

ICUnet++: an Inception-CBAM network based on Unet++ for MR spine image segmentation

Lei Li[1] · Juan Qin[1] · Lianrong Lv[1] · Mengdan Cheng[1] · Biao Wang[1] · Dan Xia[1] · Shike Wang[1]

Received: 1 November 2022 / Accepted: 4 May 2023 / Published online: 24 May 2023

Abstract
In recent years, more attention paid to the spine caused by related diseases, spinal parsing (the multi-class segmentation of vertebrae and intervertebral disc) is an important part of the diagnosis and treatment of various spinal diseases. The more accurate the segmentation of medical images, the more convenient and quick the clinicians can evaluate and diagnose spinal diseases. Traditional medical image segmentation is often time consuming and energy consuming. In this paper, an efficient and novel automatic segmentation network model for MR spine images is designed. The proposed Inception-CBAM Unet++ (ICUnet++) model replaces the initial module with the Inception structure in the encoder-decoder stage base on Unet++ , which uses the parallel connection of multiple convolution kernels to obtain the features of different receptive fields during in the feature extraction. According to the characteristics of the attention mechanism, Attention Gate module and CBAM module are used in the network to make the attention coefficient highlight the characteristics of the local area. To evaluate the segmentation performance of network model, four evaluation metrics, namely intersection over union (IoU), dice similarity coefficient(DSC), true positive rate(TPR), positive predictive value(PPV) are used in the study. The published SpineSagT2Wdataset3 spinal MRI dataset is used during the experiments. In the experiment results, IoU reaches 83.16%, DSC is 90.32%, TPR is 90.40%, and PPV is 90.52%. It can be seen that the segmentation indicators have been significantly improved, which reflects the effectiveness of the model.

Keywords Deep learning · Convolutional neural network · MRI · Spine segmentation · Attention mechanism

1 Introduction

As an important part of the human body, the spine is the axial skeleton of the human body and the pillar of the human body. In recent years, various of diseases caused by spinal cord are very common in today's world, affecting about 80% of the world's population. The consequences are not only physical pain and economic loss, but even a large number of people are disabled due to spinal cord diseases [1]. As the global population ages, spinal disorders are expected to show a significant increase in the next decade [2]. Medical image technology is commonly used in the treatment and diagnosis of spine-related diseases. Medical image works by interacting with the body in ways like X-rays, electromagnetic fields, and ultrasound, the body's internal tissues and organs morphology, density, function be expressed in the form of images, so that doctors can make health judgments based on their own knowledge and experience of the information provided in medical images. Magnetic resonance (MR) imaging and computed tomography (CT) are two common computer imaging techniques. Compared with CT technology, MR technology has clear imaging, no radiation damage, and no bone artifacts [3, 4]. MR image segmentation has become a better method for the prevention and diagnosis of spinal diseases. At first, the vertebral body is divided manually by doctors. However, this method is very time-consuming and laborious.

In recent years, researchers have paid more and more attention to the segmentation of intervertebral discs in MR spinal images. The number of related research projects have also increased. Michopoulou et al. [5] proposed to use probabilistic atlas of IVD for atlas-based segmentation. Ayed et al. [6] studied the graph cutting algorithm to segment

✉ Juan Qin
jane.qin@tjut.edu.cn

Lianrong Lv
lvlianrong@email.tjut.edu.cn

[1] School of Integrated Circuit Science and Engineering, Tianjin University of Technology, Tianjin 300384, China

the intervertebral disc. Law et al. [7] proposed an approach for intervertebral disc detection and segmentation using anisotropic directional fluxe. Rabia et al. [8] innovated a 3D intervertebral disc segmentation algorithm that that exploits weak shape priors encoded in simplex mesh active surface models. Although the above methods yielded segmentation results, they still encountered challenges and limitations of disc segmentation in MR spine images, such as distortion and rotation of the object shape, low contrast between the object and its surroundings resulting in unclear boundaries, and non-uniform intensities within the object.

Nowadays, machine learning has made significant development, especially, image segmentation based on deep learning has become a common and effective method. Convolutional neural networks (CNN) can extract features in images and perform classification, segmentation and recognition based on the obtained features [9–12]. Long et al. [13] proposed a fully convolutional network (FCN), which is the first end-to-end image semantic segmentation network for pixel-level prediction. It takes an image of any size as input, and after a series of convolution operations, its output is a high-resolution segmentation mask of the same size as the input image. Ronneberger et al. [14] proposes a U-shaped network Unet with symmetric structure, which is also composed of pure convolution. Unet has two symmetric paths, one is the encoder and the other is the decoder. In addition, skip connection is used for feature fusion between the encoder and decoder. Unet provides high-resolution feature mapping for decoder blocks, which makes the segmentation accuracy of medical images to reach a high level. Zhou et al. [15] proposed a whole new medical segmentation network based on Unet, called Unet++ . Because Unet imposes an restrictive fusion scheme on skip connection, it forces fusion only on feature graphs of the same proportion of the encoder and decoder subnets. Unet++ alleviates this problem by redesigning skip connections to aggregate different semantic scales on decoder subnetworks, thus, a more sensitive feature fusion scheme is produced and the performance of network segmentation is improved. Machine learning and deep learning methods can resolve multifaceted complications by gaining insight knowledge from simple representations. Yogesh H. Bhosale et al. [16] retrieved up to 64 published works related to deep learning-based Covid-19 detection systems for comparative analysis and discuss the challenges faced in current development. It mainly provides directions for future research to further develop effective and reliable Covid-19 detection systems. In deep learning detection systems, appropriate parameter tuning facilitates fast model tuning. Yogesh H. Bhosale et al. [17] also proposed an SSE strategy with the awareness of varied class-level accuracies for different DL models. SSE models achieve superior performance by minimizing the variance of prediction errors to the competing base learners. Several hyperparameters were also studied for the optimization model, including batch size, early stopping, epochs, and optimization strategies. There are some difficulties in the segmentation of MR spine images, such as the unclear edges between the spine and surrounding soft tissues in the image. In addition, MR image low contrast, noise, artifact and local volume effect usually reduce the performance of spine segmentation. Therefore, the image segmentation of the spine has always been a very difficult task.

For the segmentation problem of occlusion and unclear vertebral body edges in MR spine images, this paper proposes a new Inception-CBAM Unet++ (ICUnet++) network model for spine image segmentation to achieve more prominent segmentation performance. The main contribution of this paper can be divided into the following three points:

(1) Replace the convolutional layer of Unet++ encoder-decoder with Inception stucture. Inception increases the receptive field of convolutional through parallel convolutional connections, enabling ICUnet++ to obtain different scale features in the feature extraction stage and improving the segmentation ability of the network.
(2) Add Attention Gate (AG) module before each skip connection in the network. The attention mechanism can extract accurate shallow features more effectively. It helps the network more accurately locate the edges of the spine.
(3) The CBAM module joins into the ICUnet++ network. It can effectively capture region of interest(ROI) features and suppress non-ROI features, so as to strengthen the edge feature extraction ability of the network.

After the introduction, the main content of the second section is the description of related work, and the third section gives the proposed network model. The fourth section is a detailed experimental description. The fifth section is the discussion of experimental results. The sixth section is the conclusion.

2 Related work

In the past, medical image segmentation often uses handmade features for segmentation [18–20]. With the rapid progress of DCNN, in particular, Unet with its encoder-decoder framework has come to the fore, revolutionizing deep semantic segmentation of medical images. SPRNet [21] used the convolution with dilation rates of different sizes to realize the fusion of multi-scale inputs to enhance the receptive field. Fu et al. [22] used the method of average

pooling to carry out down-sampling operation on images and constructed multi-scale inputs to achieve image segmentation. Fu et al. [23] used shallow deconvolution to carry out layer upon layer superposition to extract features to the maximum extent and fully preserve position positioning. Alex Krizhevsky et al. [24] explored CNN more extensively, increased its depth and width, and greatly improved the network performance. AlexNet used two GPUs for computing, which greatly improved the compute efficiency. Since Alexnet, the breakthrough direction of convolutional neural network is to expand the depth and width of the network. However, the increase of the depth and width of the network will lead to a sharp increase in parameters, resulting in overfitting and higher computational complexity. On the other hand, the deeper the network, the more prone it is to gradient disappearance (gradient dispersion), making it difficult to train and optimize the model. Inception [25] was born out of such circumstances. The Inception model has two main advantages. First of all, 1×1 convolution is used to carry out lifting and lowering dimensions, by stacking more convolution in modules of the same size, richer features can be extracted. The second is convolution reaggregation on multiple dimensions simultaneously. Intuitively, convolution at multiple scales can extract features of different scales, richer features also mean that the final classification judgment is more accurate. Inspired by the Inception-ResNet model, Gu et al. [26] designed a novel dense atrous convolution (DAC), which used the 33 continuous convolution layers and pooling layers to capture multi-scale features within a limited scale range in the coding stage. Rad et al. [27] conducted a study in the multi-scale directions and found that a larger receptive field could obtain information at any position in the input image. And Zhang et al. [28] designed an Inception-RES module in the network, which can not only fuse multi-scale features but also solve the gradient problem through residual connection. Li et al. [29] designed a novel Dilated-Inception convolutional network to extract and locate features in images. Oktay et al. [30] proposed an Attention U-net structure with an attention mechanism to segment the pancreas. Xiao et al. [31] designed a convolutional network with a weighted attention mechanism and added skip connections to segment high-resolution retinal vessels. Zhang et al. [32] designed a network structure called AG-Net with "attention-guided filter", which has an excellent function for preserving feature information. Guo et al. [33] designed the lightweight SAU-Net model based on AG-Net, which can largely eliminate the overfitting problem by using the special mechanism of DropBlock [34]. SAU-Net can be well trained even with small sample datasets. SeNet proposed by Hu et al. [35] is also designed by using the characteristics of attention in the network. The squeeze-and-excitation structure is that the contraction operation for the features generated after convolution gets the overall feature information at the channel level, and then the excitation operation is applied. The network learns the nonlinear relationship between each channel and obtains the feature weights of different channels. The advantage of this attention mechanism is that it can make the network prefer the relevant channel information in the training process, while suppressing the irrelevant information. However, its disadvantage lies in that it only focuses on the relationship between feature channels and lacks the information linkage between contexts. Woo et al. [36] proposed CBAM. Compared with SeNet structure, CBAM adds spatial attention after channel attention. The max-pooling operation is added after the channel attention in SeNet structure, aiming to take the features extracted from the channel attention as the input of the following spatial attention module. This method saves the number of parameters and computational complexity, and brings a more powerful improvement in network performance. Table 1 summarizes the advantages and disadvantages of some existing networks in this section.

According to the relevant work mentioned above, this paper proposed an ICUnet++ model with a redesigned Inception-CBAM for MR spine image segmentation. In the proposed ICUnet++ , the AG and CBAM block have been added to better focus on edge feature information. In addition, ICUnet++ uses Inception structure to supersede the two layers of 3×3 convolution in the original network, which aims to increase the receptive field by using the multi-scale feature fusion method to achieve performance improvement.

3 Method

This section mainly introduces the proposed ICUnet++ . It elaborates the proposed ICUnet++ network model, Inception structure, and CBAM module respectively.

3.1 Network model framework

The proposed ICUnet++ network model for spine image segmentation is derived from the encoder-decoder architecture Unet++ model with dense skip connections. As shown in Fig. 1, the network replaces the two layers 3×3 convolution in the encoder-decoder stage with the designed Inception structure. The Inception structure uses multiple convolutional kernels of different sizes to obtain features at different scales and enhance the feature extraction capability of the network. The network uses a 2×2 maximum pooling for downsampling, doubling the number of feature channels downsampled at each layer. Accordingly, the upsampling is required in the decoder stage to restore the features extracted in the encoding stage to their original size, and

Table 1 Advantages and disadvantages of network models in related work

Model	Advantages	Disadvantages
SPRNet [21]	SPRNet achieves efficient instance segmentation by introducing a single-pixel reconstruction (SPR) branch in an off-the-shelf single-stage detector	Despite using atrous convolution to compact spatial information into a single pixel, it is still challenging to recover a very detailed and accurate mask from pixels
M-Net [22]	M-Net jointly solves the optic discand optic cup partitioning problem in a single-stage multi-label system	Optic cup boundary cannot be recognized in the case of blurred image and low contrast
AlexNet [24]	AlexNet increases network depth and width, greatly improving network performance. The computing efficiency is greatly improved by using two GPUs simultaneously	AlexNet's asks longer periods of time for training and matching the current path of the human visual system
CENet [26]	CENet proposed a dense atrous convolution block and a residual multicore pool block to capture more advanced features and retain more spatial information	CENet has only been validated on two-dimensional images so far, and has yet to be extended on three-dimensional data
DIUNet [28]	DIUNet is deeper and wider network. Intensive connection is adopted to avoid gradient disappearing or redundant calculation during network training	Too many parameters make model training more difficult and slow
Attention U-net [30]	Attention Gates can be easily integrated into the standard CNN architecture with minimal computational overhead while increasing model sensitivity and prediction accuracy	Attention U-net experiments using residual connections did not provide any significant performance improvement
SAUNet [33]	SA-UNet introduces a spatial attention module and multiplies the attention map by the input feature map for adaptive feature refinement	SAUNet only paid attention to spatial attention and did not conduct corresponding experiments on channel attention

up-sampling is performed by means of adjacent interpolation. During the upsampling process, the number of feature channels is halved. In order to obtain more useful global information, CBAM module is added in the decoder stage of the network. In the skip connection stage, dense nested connections are used to increase the depth and width of the network and integrate image features at different levels. The attention mechanism is used to focus on relevant information and ignore irrelevant information, so AG and CBAM are added to the skip connection of ICUnet++ .

3.2 Inception structure

ICUnet++ network model is encoder-decoder architecture. The encoder stage tries to capture more high-level semantic features in the input image while gradually reducing its spatial dimensionality. And the decoder stage is to recover the original resolution and spatial dimensionality of the image. To obtain more contextual information, features can be extracted in the encoder-decoder using convolutional kernels of different scales. Therefore, Inception is used to replace the convolution block in the encoder-decoder in Unet++. The Inception structure is shown in Fig. 2, where (a) is the initial structure, and (b) is modified structure. The structure replaces Max_pool in the original module with 3×3 convolution and the number of input channels remains constant. The stride is set to 1 in the network, which increases the number of extracted features. Then batch normalization (BN) [37] is added after each convolution to reduce the gradient disappearance or explosion. In the proposed network, two 3×3 convolutions in Unet++ are replaced by the designed Inception block. Compared with the original Unet++ structure, ICUnet++ obtains features at different scales from convolutional kernels of different sizes, increasing the receptive field for feature extraction. Define that y_l is the l_{th} layer output. The $h_{n\times n}()$ denotes a $n\times n$ convolutional layer and $h_b()$ represents the BN layer. And $f_r()$ denotes the ReLU activation function, concatenation function is denoted by $\circ$. The output of each modified Inception structure can be expressed as formula (1):

$$\begin{aligned} y_{l+1} &= f_r\left[(h_b(h_{3\times3}(y_l)))\right] \\ \circ f_r\big[h_b(h_{3\times3}(f_r(h_b(h_{3\times3}(f_r(h_b(h_{1\times1}(y_l)))))))\big] \\ &\circ f_r\left[(h_b(h_{3\times3}(y_l)))\right] \end{aligned} \tag{1}$$

3.3 CBAM and AG module

Convolution Block Attention Module (CBAM) contains 2 independent submodules, Channel Attention module and Spatial Attention module. As a lightweight general structure, CBAM is used in feedforward convolutional networks, which

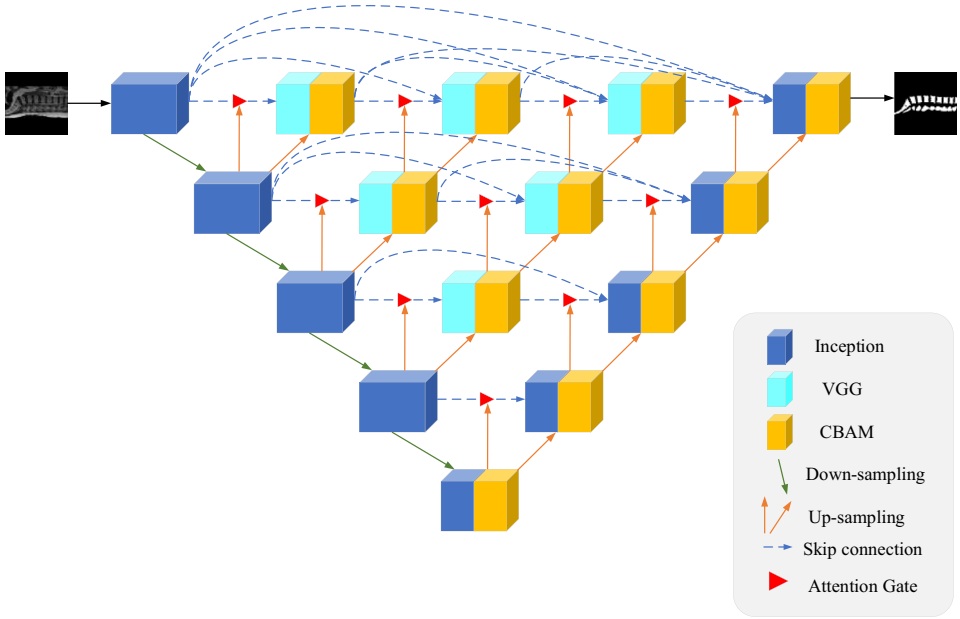

Fig. 1 The proposed ICUnet++ model

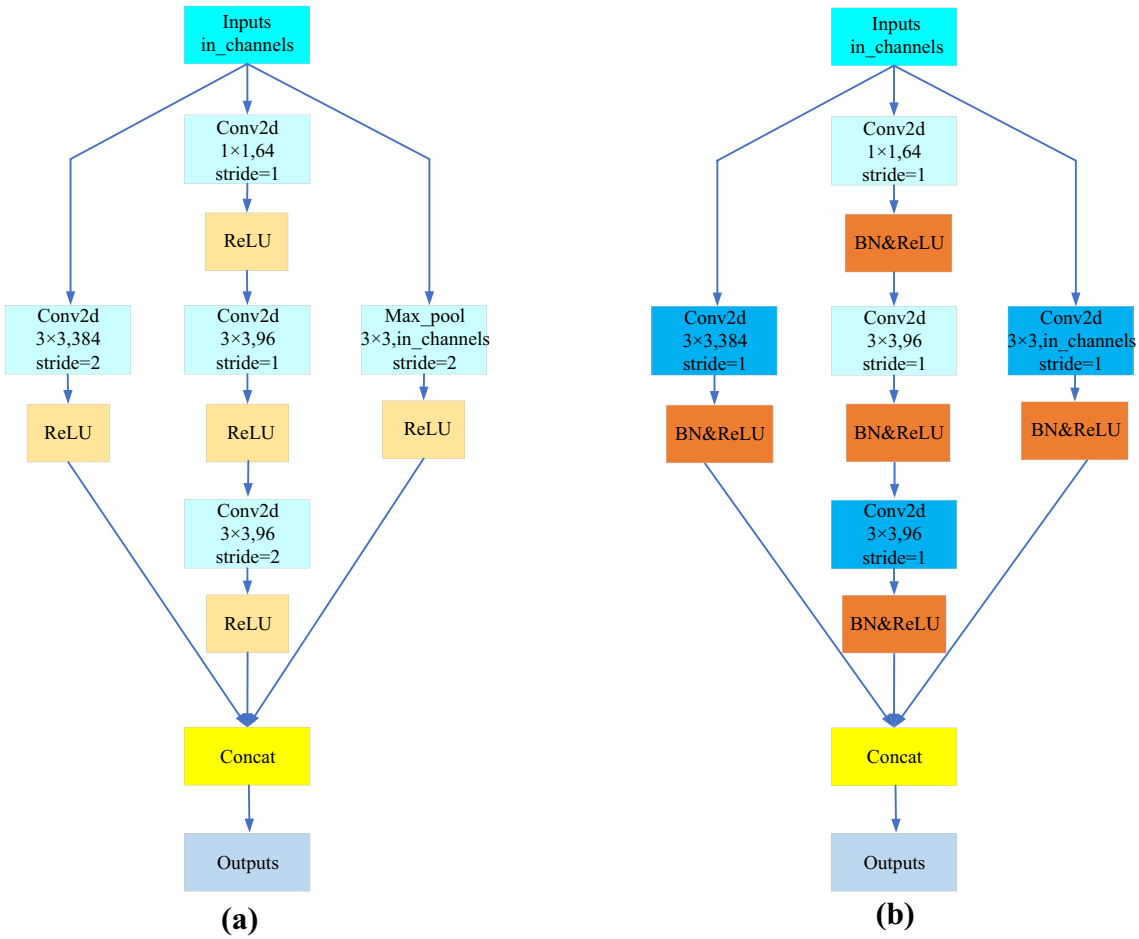

Fig. 2 Inception structure **a** Initial structure, **b** Modified structure

can be seamlessly integrated into other CNN networks, and it brings a negligible number of parameters.

The CBAM module inputs the intermediate feature map, first calculates a one-dimensional channel attention map, multiplies it with the intermediate feature map, then calculates a 2D spatial attention map, and multiplies it with the feature map of the previous layer for adaptive feature refinement. During multiplication, the attention values are broadcasted accordingly: channel attention values are broadcasted along the spatial dimension, and vice versa. The.

CBAM is shown in Fig. 3. The calculation process of CBAM is expressed as follows:

$$\begin{cases} \mathbf{F}' = \mathbf{M_c}(\mathbf{F}) \otimes \mathbf{F} \\ \mathbf{F}'' = \mathbf{M_s}(\mathbf{F}') \otimes \mathbf{F}' \end{cases} \tag{2}$$

where $\mathbf{F} \in R^{C\times H\times W}$ is the intermediate feature map, $\mathbf{M_C} \in R^{C\times 1\times 1}$ calculates the one-dimensional channel attention map, and $\mathbf{M_s} \in R^{1\times H\times W}$ calculates the 2D spatial attention map. The $\otimes$ denotes element-wise multiplication. $\mathbf{F}''$ is the refined feature map.

Figure 4 shows the AG module. It's inputs are the upsampled features in the extended path and the corresponding features of the encoder. The former is used as a gating signal to enhance the learning of target regions relevant for the segmentation task, while suppressing task-irrelevant regions. Thus, attention gating can improve the efficiency of semantic information propagation through skip connections. The s-shaped activation function sigmoid is chosen to train the convergence of the parameters within the gate and to obtain the attention coefficient α_i. The refined features are obtained by multiplying the encoder features by the coefficient α_i.

3.4 Loss function

The loss function in deep learning is a way to evaluate the gap between the actual value and the predicted value of the neural network output. The network selects the appropriate loss function to help improve the accuracy of image segmentation. The network uses the combined loss function of dice loss and binary cross-entropy loss, and applies it to the output of each different level. The expressions are shown in formula (3):

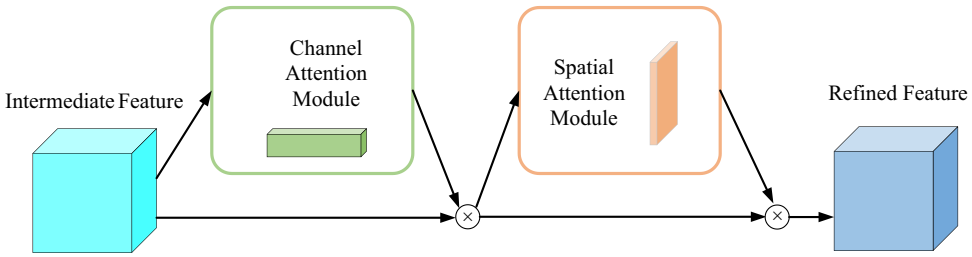

Fig. 3 Convolution Block Attention Module

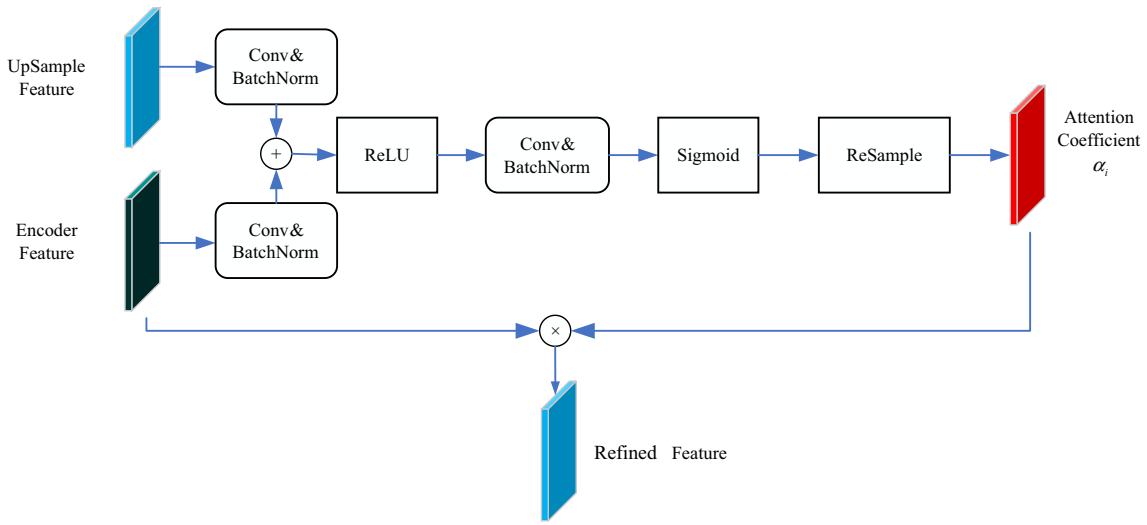

Fig. 4 Attention Gate module

$$L(Y,\hat{Y}) = -\frac{1}{N}\sum_{b=1}^{N}\frac{1}{2}\left(\log \hat{Y}_b + \frac{2\cdot Y_b \cdot \hat{Y}_b}{Y_b + \hat{Y}_b}\right) \tag{3}$$

where $\hat{Y}_b$ represents the prediction probability and Y_b represents the basic authenticity of the image. N indicates the batch size.

4 Experiments and results

In this section, the data set, evaluation metrics, experimental and results are presented in detail to demonstrate the reliability of this model.

4.1 Experimental Data

The publicly available SpineSagT2WDataset3 is used to train and evaluate the segmentation performance of ICUnet++ . All data are collected with the same equipment, and the magnetic field strength is 3.0 T. The dataset contains sagittal T2-weighted volume data from 195 spinal patients, such as lumbar disc herniation and lumbar disc degeneration. The ground truth (GT) of the dataset is manually drawn by the medical imaging expert with a vertebral label of 1 and a background label of 0. The 3D volume data are converted into 2460 slice 2D images of 880×880 pixels. Each spinal slice 2D image contains at least eight vertebrae. The preprocessing operation is to adjust the image pixels, and the pixel size after adjustment is 256×256.

4.2 Evaluation metrics

In this study, intersection over union (IoU), dice similarity coefficient (DSC), true positive rate (TPR), and positive predictive value (PPV) are selected to evaluate the performance of the network model. IoU and DSC are used to evaluate the similarity. TPR describes the proportion of true positive samples to all positive samples. The larger the TPR indicates the less missed identification, and the smaller the TPR indicates the more missed identification. PPV is the proportion of true positive samples to all predicted positive samples. The larger PPV indicates fewer false detections, and the smaller PPV indicates more false detections. The above four metrics are defined as follows.

$$\mathrm{DSC} = \frac{2\left|R_{gt} \bigcap R_{pred}\right|}{\left|R_{gt}\right| + \left|R_{pred}\right|} = \frac{2\mathrm{TP}}{2\mathrm{TP} + \mathrm{FP} + \mathrm{FN}} \tag{4}$$

$$\mathrm{IoU} = \frac{\left|R_{gt} \bigcap R_{pred}\right|}{\left|R_{gt} \bigcup R_{pred}\right|} = \frac{\mathrm{TP}}{\mathrm{TP} + \mathrm{FP} + \mathrm{FN}} \tag{5}$$

$$\mathrm{TPR} = \frac{R_{gt} \bigcap R_{pred}}{R_{gt}} = \frac{\mathrm{TP}}{\mathrm{TP} + \mathrm{FN}} \tag{6}$$

$$\mathrm{PPV} = \frac{R_{gt} \bigcap R_{pred}}{R_{pred}} = \frac{\mathrm{TP}}{\mathrm{TP} + \mathrm{FP}} \tag{7}$$

where TP, FP and FN denote true positives, false positives and false negatives, respectively. R_{gt} and R_{pred} are ground truth and predicted segmentation results respectively.

4.3 Training details

During model training, the dataset is divided in 4:1 ratio. 80% of the sliced 2D images are used for training the network and adjusting the network parameters, and the remaining 20% are used to evaluate the performance of the model. Each patient's MRI data consiste of 12 to 18 binary slices, with a total of 2460 slices for 195 patients, of which 1968 are used for training and 492 for validation. Figure 5 shows a set of MRI slice images of a patient and the corresponding GT in dataset SpineSagT2WDataset3.

In addition, the code is implemented in Python 3.8 and Pytorch 1.12.0 enviroment. The objective function is optimized using the Adam optimizer. The learning rate is 0.001. The batch size is 3. The training epochs are 150 in total. The

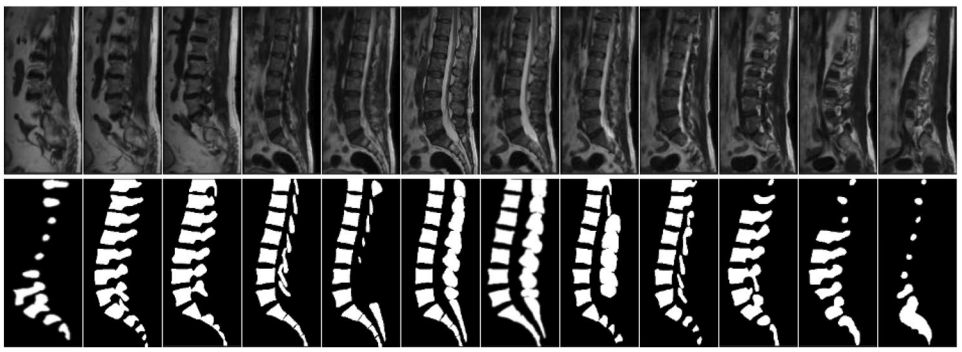

Fig. 5 The 12 slices and the corresponding GTs converted from MRI volume data

proposed model is trained and evaluated on a workstation with NVIDIA GeForce RTX 3090 GPU.

4.4 Experimental results

This experiment improves the performance by modifying the basic module of Unet++ and adding the attention module. Four evaluation metrics, IoU, DSC, TPR, and PPV are used to evaluate the performance of the model. All experiments in this study were repeated three times to demonstrate the stability of the model, and the best value was taken as the experimental result. Finally, the mean and standard deviation of the proposed ICUnet++ and the other network model is calculated to show the stability of the model. First, five sets of comparative experiments are conducted in this paper based on Unet++ to demonstrate the effect of the attention position on performance. (1). Put the spine dataset into the original Unet++ for training. (2). Add CBAM in skip connection stage, named Unet++ C1. (3). Add CBAM in the encoder stage and skip connection stage, named Unet++ C2. (4). Add CBAM in the decoder stage and skip connection stage, named Unet++ C3. (5). Add CBAM in the encoder-decoder and skip connection stage, named Unet++ C4. (6). Add CBAM in decoder stage, and add AG module and CBAM in skip connection stage, named Unet++ C5. As shown in Table 2, adding the attention mechanism to the Unet++ network can effectively improve the performance of spine image segmentation. IoU, DSC, and TPR achieve the best results in Unet++ C5. with 82.03%, 89.50%, and 90.00%, respectively. Figure 6 shows the segmentation results and error plots of six different Unet++ models.

Further experiment attempts are made to modify the encoder-decoder of Unet++. The modified Inception structure is used to improve the segmentation performance of the network. A total of 4 groups of related experiments have been done. (1). Replace the convolutional layer of Unet++ encoder-decoder with Inception structure, and add AG module and CBAM in skip connection, named

Table 2 Comparison of IoU, DSC, TPR, and PPV of different Unet++ models

Model	Encoder	Decoder	Skip-connection		IoU	DSC	TPR	PPV
	CBAM	CBAM	AG	CBAM				
Unet++					0.8170	0.8925	0.8913	0.8927
Unet++ C1				√	0.8177	0.8933	0.8925	0.8978
Unet++ C2	√			√	0.8168	0.8924	0.8929	0.8940
Unet++ C3		√		√	0.8188	0.8936	0.8921	0.8962
Unet++ C4	√	√		√	0.8195	0.8945	0.8922	**0.8988**
Unet++ C5		√	√	√	**0.8203**	**0.8950**	**0.9000**	0.8921

Bold values represent the best value for each evaluation indicator in each table

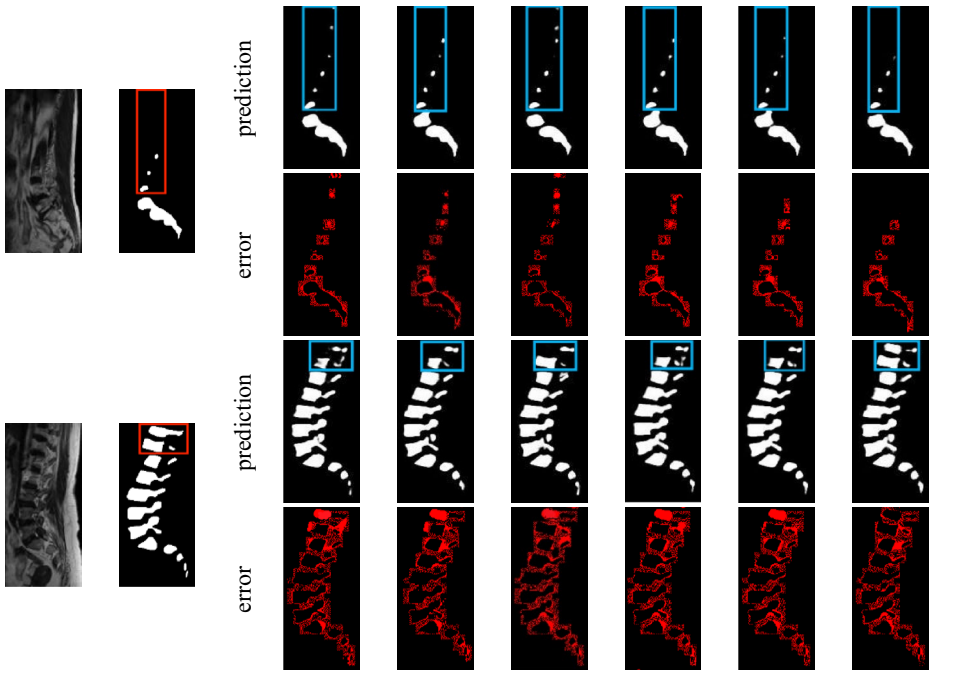

Fig. 6 Segmentation results of 6 different Unet++ models

Table 3 Comparison of IoU, DSC, TPR, and PPV of 5 different ICUnet++ models

Model	Encoder	Decoder	Skip-connection		IoU	DSC	TPR	PPV
	CBAM	CBAM	AG	CBAM				
ICUnet++1			√	√	0.8227	0.8973	0.8972	0.8988
ICUnet++2	√		√	√	0.8216	0.8963	0.8954	0.8991
ICUnet++3	√	√	√	√	0.8268	0.8989	0.9028	0.8988
ICUnet++(Ours)		√	√	√	**0.8316**	**0.9032**	**0.9040**	**0.9052**

Bold values represent the best value for eachevaluation indicator in each table

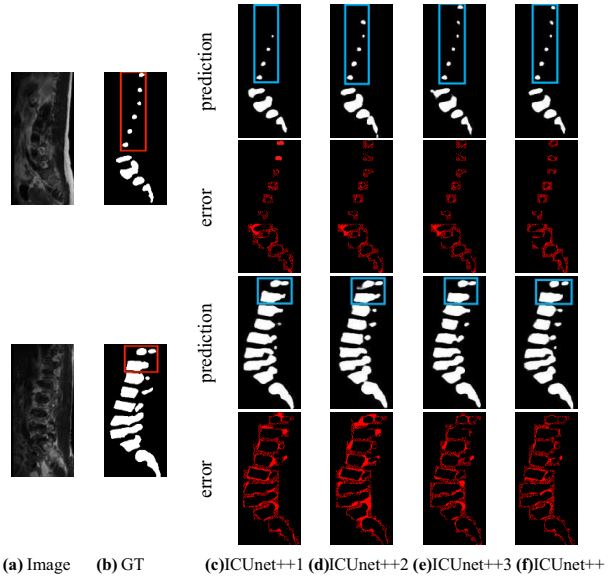

Fig. 7 Segmentation results of 4 different ICUnet++ models

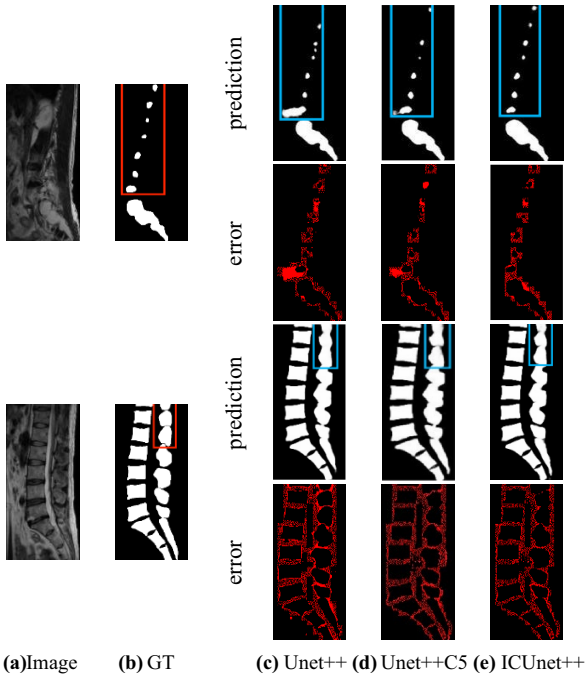

Fig. 8 Comparison of Unet++, Unet++C5, and ICUnet++ segmentation

Table 4 Comparison of Unet, Unet++, ResUnet, DenseUnet, RAR-Unet, mRR-Unet, and ICUnet++

Model	IoU	DSC	TPR	PPV	Time(s)/epoch
Unet[14]	0.8142	0.8920	0.8925	0.8916	25 s
Unet++[15]	0.8170	0.8925	0.8913	0.8927	40 s
ResUnet[38]	0.8169	0.8930	0.8956	0.8926	32 s
DenseUnet[39]	0.8137	0.8909	0.8913	0.8928	37 s
RARUnet[40]	0.8215	0.8950	0.8972	0.8946	43 s
mRRUnet[41]	0.8212	0.8945	0.8977	0.8950	50 s
ICUnet++(Ours)	**0.8316**	**0.9032**	**0.9040**	**0.9052**	152 s

Bold values represent the best value for eachevaluation indicator in each table

ICUnet++ 1. (2). Replace the convolutional layer of Unet++ encoder-decoder with the Inception structure, add CBAM after the Inception structure in the encoder stage, and add AG module and CBAM to skip connection, named ICUnet++ 2. (3). Replace the convolutional layer of encoder-decoder of Unet++ with Inception structure, add CBAM after Inception of encoder-decoder stage, and add AG module and CBAM in skip connection, named ICUnet++ 3. (4). Replace the convolutional layer of Unet++ encoder-decoder with Inception structure, add CBAM after Inception in the decoder stage, and add AG module and CBAM in the skip connection, named ICUnet++ . Table 3 shows the experimental data comparison. ICUnet++ has achieved the best results. IoU, DSC, TPR, and PPV, are 83.16%, 90.32%, 90.40%, and 90.52% respectively. Figure 7 shows the segmentation results and error plots of these four models. As shown by experimental data, the segmentation performance has been effectively improved after adding the Inception structure. Figure 8 shows the comparison of Unet++, Unet++C5 and ICUnet++ segmentation results and error plots. Figure 8d illustrates that increasing the attention can better segment the vertebral edges in MR spine images compared to Fig. 8c, but there is still the problem of missing vertebrae during the segmentation process. Figure 8e illustrates that the use of multi-scale feature extraction based on Fig. 8d makes the model obtain better segmentation performance, which is closer to the ground truth. The proposed

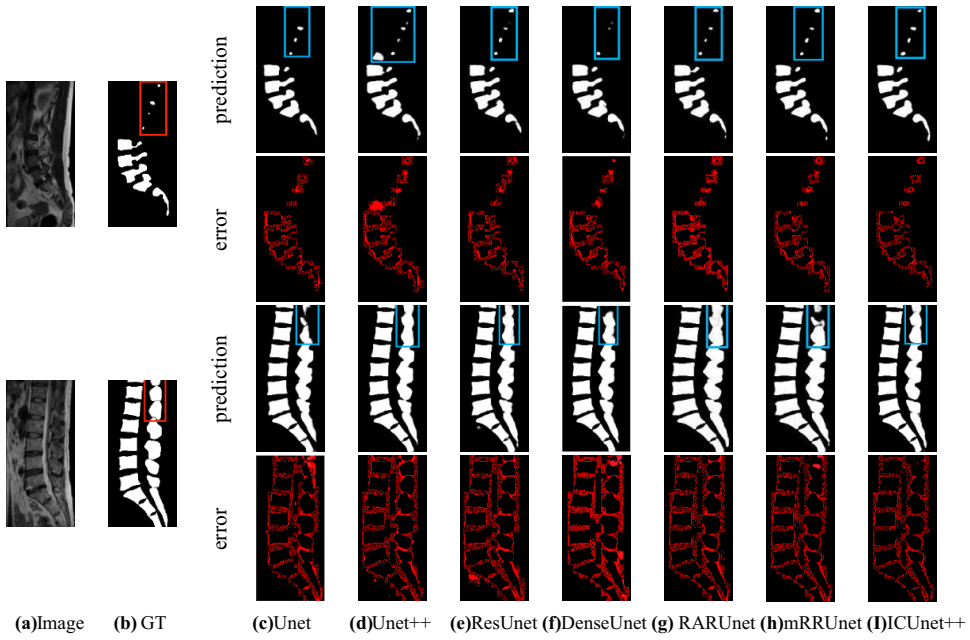

Fig. 9 The proposed ICUnet++ is compared with other five network models

Table 5 Mean and standard deviation of Unet, Unet++, ResUnet, DenseUnet, RAR-Unet, mRR-Unet, and ICUnet++

Model	IoU	DSC	TPR	PPV
Unet[14]	0.8138 + 0.00039	0.8917 + 0.00041	0.8919 + 0.00058	0.8912 + **0.00037**
Unet++ [15]	0.8165 + 0.00041	0.8921 + 0.00056	0.8922 + 0.00070	0.8920 + 0.00050
ResUnet[38]	0.8165 + 0.00037	0.8924 + 0.00069	0.8957 + 0.00070	0.8925 + 0.00041
DenseUnet[39]	0.8132 + 0.00039	0.8915 + 0.00054	0.8921 + 0.00068	0.8934 + 0.00045
RAR-Unet[40]	0.8210 + 0.00045	0.8945 + 0.00058	0.8966 + 0.00059	0.8945 + 0.00086
mRR-Unet[41]	0.8207 + 0.00037	0.8950 + 0.00056	0.8983 + 0.00058	0.8957 + 0.00057
ICUnet++ (Ours)	**0.8313 + 0.00029**	**0.9028 + 0.00040**	**0.9035 + 0.00041**	**0.9046** + 0.00049

Bold values represent the best value for each evaluation indicator in each table

Table 6 Metric comparison for different dataset divisions

Dataset for different divisions	IoU	DSC	TPR	PPV
ICUnet++ a	0.8304	0.9013	0.9035	0.9041
ICUnet++ b	0.8285	0.9001	0.8999	0.9004
ICUnet++ c	0.8279	0.8994	0.8984	0.8995
ICUnet++ d	0.8310	0.9020	0.9028	0.9033
ICUnet++	0.8316	0.9032	0.9040	0.9052

ICUnet++ can better segment MR spine image occlusion and unclear vertebral body edges. Table 4 shows the experimental results comparison of the ICUnet++ and other five network models on SpineSagT2WDataset3, where ICUnet++ has better segmentation performance. Figure 9 shows the segmentation results and error plots of Unet, Unet++, ResUnet, DenseUnet, RAR-Unet, and ICUnet++ , it can be seen that ICUnet++ has better segmentation performance. In order to demonstrate the stability and robustness of the ICUnet++ model, relevant experiments are done in this study. Table 5 shows the mean and standard deviation of each metric for ICUnet++ and the other network structure. It can be seen that ICUnet++ achieves better values in IoU, DSC, and TPR, while only the standard deviation of PPV is higher than that of Unet. To explain the robustness of the model, the open dataset is randomly partitioned so that each training and validation images are different, simulating a different data source. Five separate training and validation experiments were conducted, namely the corresponding ICUnet++ a, ICUnet++ b, ICUnet++ c, ICUnet++ d, and ICUnet++ . Table 6 shows the experimental results to illustrate the robustness of ICUnet++ .

5 Discussion

The results of spine medical image analysis serve as an important clinical indicator that helps doctors to better diagnose and treat patients. Since the traditional Unet and Unet++ are performed on a relatively single convolution kernel, there are limitations in feature extraction

at different scales. Therefore, based on multi-scale convolution and attention mechanism, the ICUnet++ for automatic segmentation of MR spine images is proposed. The modified Inception structure allows the network to obtain multi-scale feature information and perform fusion. In addition, according to the characteristics of attention mechanism, AG module and CBAM are added to make the network more focused on the extraction of important features. In Tables 2 and 3, the same conclusion can be drawn that there are differences in the result of adding CBAM in different locations. For both Unet++ and ICUnet++ , adding CBAM only in the encoder stage lead to the reduction of segmentation performance. On the contrary, adding CBAM only in the decoder stage will have a better effect, this is due to the fact that the spatial feature map at the beginning of the encoder stage is too large and the number of channels is too small. The extracted channel weights are too generalized without falling into some specific features, and the extracted spatial weights is not generalized enough due to the small number of channels. The spatial attention is sensitive and difficult to learn, which is more likely to cause negative effects. The designed ICUnet++ model uses the Inception structure in the encoder-decoder stage, adds CBAM after the convolution in the decoding stage, and adds the AG module and CBAM in the skip connection stage. Figure 8 shows that the segmentation of Unet++ C5 at the vertebral edges is better compared to Unet++ after adding attention. The segmentation performance of ICUnet++ is even better after replacing the convolutional layer in the encoding–decoding stage of Unet++ C5 using the Inception structure. And according to the experimental data in Table 4 and the segmentation maps in Fig. 9, it can be seen that ICUnet++ has better segmentation performance compared to Unet, Unet++, ResUnet, DenseUnet, RAR-Unet and mRR-Unet. ICUnet++ is trained and validated on the public spine dataset SpineSagT2Wdataset3. The experimental results show that ICUnet++ has better segmentation performance compared to the original network and the other network models, which is helpful for automatic analysis and intelligent diagnosis of spine MRI images.

Furthermore, due to the limitations of the dataset, this study still needs to be optimized. In future research, the main task is to find more medical image datasets to validate the proposed model. ICUnet++ has a good optimization in model performance, however, the training time is relatively long as shown in Table 4. So a lightweight model will be designed in future work.

6 Conclusion

Various diseases caused by the spine have a significant negative impact on our daily life, and the inaccurate segmentation of MR spine images with occlusion and unclear vertebral body edges can easily lead to medical misdiagnosis with serious consequences. Therefore, this study proposes the ICUnet++ model to implement the segmentation of spine images. The attention mechanism is introduced into the ICUnet++ model, which enhances the feature extraction for the edge details of spine images. ICUnet++ also introduces the modified Inception structure into Unet++ to replace the original VGG module. It uses multi-scale feature extraction of spine images and improve the model segmentation performance. The robustness of the model is experimentally verified by randomly dividing the dataset and simulating different data sources. The comparison of various experimental data with other existing network models proves that ICUnet++ has better segmentation performance, which is beneficial for automatic analysis of spine images and intelligent diagnosis of spine diseases.

Acknowledgements The authors acknowledge support from the Tianjin municipal education commission scientific research project, China (Grant No. 2018KJ132) and Tianjin Research Innovation Project for Postgraduate Students (No. 2022SKYZ255) Correct project name: Tianjin Research Innovation Project for Postgraduate Students Supplementary project number: 2022SKYZ255.

Data availability The data that support the conclusions of this study are openly available in SpineSagT2Wdataset3 Dataset at https://pan.baidu.com/s/1_N9v9UWWArPbq3h0oqhZ5Q.

Declarations

Conflict of Interest We declare that we have no conflict of interest.

Ethical approval This article does not contain any studies with human participants performed by any of the authors.

References

1. Freburger JK, Holmes GM, Agans RP et al (2009) The rising prevalence of chronic low back pain. Arch Intern Med 169(3):251–258. https://doi.org/10.1001/archinternmed.2008.543
2. Woolf AD, Pleger B (2003) Burden of major musculoskeletal conditions. Bull World Health Organ. 81(9):646–56. https://pubmed.ncbi.nlm.nih.gov/14710506/. Accessed 1 Oct 2022
3. Emch TM, Modic MT (2011) Imaging of lumbar degenerative disk disease: history and current state. Skeletal Radiol 40:1175–1189. https://doi.org/10.1007/s00256-011-1163-x
4. Li S, Liu J, Song Z (2022) Brain tumor segmentation based on region of interest-aided localization and segmentation U-Net. Int J Mach Learn Cyber 13:2435–2445. https://doi.org/10.1007/s13042-022-01536-4

5. Michopoulou SK, Costaridou L, Panagiotopoulos E, Speller R, Panayiotakis G, Todd-Pokropek A (2009) Atlas-based segmentation of degenerated lumbar intervertebral discs from MR images of the spine. IEEE Trans Biomed Eng 56(9):2225–2231. https://doi.org/10.1109/tbme.2009.2019765
6. Ben Ayed I, Punithakumar K, Garvin G, Romano W, Li S (2011) Graph Cuts with Invariant Object-Interaction Priors: In: Székely G, Hahn HK (eds) Information Processing in Medical Imaging. IPMI 2011. Lecture Notes in Computer Science, vol 6801. https://doi.org/10.1007/978-3-642-22092-0_19
7. Law MW, Tay K, Leung A, Garvin GJ, Li S (2013) Intervertebral disc segmentation in MR images using anisotropic oriented flux. Med Image Anal 17:43–61. https://doi.org/10.1016/j.media.2012.06.006
8. Haq R, Besachio DA, Borgie RC, Audette MA (2014) Using shape-aware models for lumbar spine intervertebral disc segmentation. In: 2014 22nd International Conference on pattern recognition, pp 3191–3196. https://doi.org/10.1109/ICPR.2014.550
9. Ma Y, Xie Y (2022) Evolutionary neural networks for deep learning: a review. Int J Mach Learn & Cyber 13:3001–3018. https://doi.org/10.1007/s13042-022-01578-8
10. Liu Q, Zhang J, Liu J et al (2022) Feature extraction and classification algorithm, which one is more essential? An experimental study on a specific task of vibration signal diagnosis. Int J Mach Learn Cyber 13:1685–1696. https://doi.org/10.1007/s13042-021-01477-4
11. Ottoni ALC, de Amorim RM, Novo MS et al (2022) Tuning of data augmentation hyperparameters in deep learning to building construction image classification with small datasets. Int J Mach Learn Cyber. https://doi.org/10.1007/s13042-022-01555-1
12. Li F, Gao D, Yang Y et al (2022) Small target deep convolution recognition algorithm based on improved YOLOv4. Int J Mach Learn Cyber. https://doi.org/10.1007/s13042-021-01496-1
13. Long J, Shelhamer E, Darrell T (2017) Fully convolutional networks for semantic segmentation. IEEE Trans Pattern Anal Mach Intell. https://doi.org/10.1109/TPAMI.2016.2572683
14. Ronneberger, O, Fischer, P, Brox T (2015) U-Net: convolutional networks for biomedical image segmentation. In: Navab N, Hornegger J, Wells W, Frangi A (eds) Medical Image Computing and Computer-Assisted Intervention—MICCAI 2015. MICCAI 2015. Lecture Notes in Computer Science, vol 9351. https://doi.org/10.1007/978-3-319-24574-4_28
15. Zhou Z, Siddiquee MMR, Tajbakhsh N, Liang J (2018) UNet++: a nested U-Net architecture for medical image segmentation. In: et al. Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support. 11045:3-11. https://doi.org/10.1007/978-3-030-00889-5_1
16. Bhosale YH, Patnaik KS (2022) Application of deep learning techniques in diagnosis of Covid-19 (Coronavirus): a systematic review. Neural Process Lett. https://doi.org/10.1007/s11063-022-11023-0
17. Bhosale YH, Patnaik KS (2023) PulDi-COVID: chronic obstructive pulmonary (lung) diseases with COVID-19 classification using ensemble deep convolutional neural network from chest X-ray images to minimize severity and mortality rates. Biomed Signal Processi Control 81:104445. https://doi.org/10.1016/j.bspc.2022.104445
18. Chen W, Smith R, Ji S-Y, Ward KR, Najarian K (2009) Automated ventricular systems segmentation in brain ct images by combining low-level segmentation and highlevel template matching. BMC Med Inf Decis Suppl 1(Suppl 1):S4. https://doi.org/10.1186/1472-6947-9-s1-s4
19. Zhu X, Rangayyan RM (2008) Detection of the optic disc in images of the retina using the Hough transform. In: 2008 30th Annual International Conference of the IEEE Engineering in Medicine and Biology Society, IEEE, pp 3546–3549. https://doi.org/10.1109/iembs.2008.4649971
20. Mihaylova A, Georgieva V (2018) Spleen segmentation in MRI sequence images using template matching and active contours. Procedia Comput Sci 131:15–22. https://doi.org/10.1016/j.procs.2018.04.180
21. Yu J, Yao J, Zhang J, Yu Z, Tao D (2020) Sprnet: single-pixel reconstruction for onestage instance segmentation. IEEE Trans Cybern 51(4):1731–1742. https://doi.org/10.1109/TCYB.2020.2969046
22. Fu H, Cheng J, Xu Y, Wong DWK, Liu J, Cao X (2018) Joint optic disc and cup segmentation based on multi-label deep network and polar transformation. IEEE Trans Med Imaging 37(7):1597–1605. https://doi.org/10.1109/TMI.2018.2791488
23. Fu J, Liu J, Wang Y, Zhou J, Wang C, Lu H (2019) Stacked deconvolutional network for semantic segmentation. IEEE Trans Image Process. https://doi.org/10.1109/TIP.2019.2895460
24. Krizhevsky A, Sutskever I, Hinton GE (2012) ImageNet classification with deep convolutional neural networks. Commun ACN 60:84–90. https://doi.org/10.1145/3065386
25. C. Szegedy et al. (2015) Going deeper with convolutions. In: 2015 IEEE Conference on computer vision and pattern recognition (CVPR), pp 1–9. https://doi.org/10.1109/CVPR.2015.7298594
26. Gu Z et al (2019) CE-Net: context encoder network for 2D medical image segmentation. IEEE Trans Med Imaging 38(10):2281–2292. https://doi.org/10.1109/TMI.2019.2903562
27. Rad RM, Saeedi P, Au J, Havelock J (2020) Trophectoderm segmentation in human embryo images via inceptioned U-Net. Med Image Anal 62:101612. https://doi.org/10.1016/j.media.2019.101612
28. Zhang Z, Wu C, Coleman S, Kerr D (2020) DENSE-INception U-net for medical image segmentation. Comput Methods Programs Biomed 192:105395. https://doi.org/10.1016/j.cmpb.2020.105395
29. Li J, Yu ZL, Gu Z, Liu H, Li Y (2019) Dilated-inception net: multi-scale feature aggregation for cardiac right ventricle segmentation. IEEE Trans Biomed Eng 66(12):3499–3508. https://doi.org/10.1109/tbme.2019.2906667
30. Oktay O, Schlemper J, Folgoc L, Lee M, Heinrich M, Misawa K, Mori K, McDonagh S, Hammerla N, Kainz B, Glocker B, Rueckert D (2018) Attention U-Net: learning where to look for the pancreas. https://doi.org/10.48550/arXiv.1804.03999
31. Xiao X, Lian S, Luo Z, Li S (2018) Weighted Res-Unet for high-quality retina vessel segmentation. In: 2018 9th International Conference on information technology in medicine and education (ITME), pp 327–331. https://doi.org/10.1109/ITME.2018.00080
32. Zhang S, Fu H, Yan Y, Zhang Y, Wu Q, Yang M, Tan M (2019) Attention guided network for retinal image segmentation. In: Medical Image Computing and Computer Assisted Intervention—MICCAI 2019. Lecture Notes in Computer Science, vol 11764. https://doi.org/10.1007/978-3-030-32239-7_88
33. Guo C, Szemenyei M, Yi Y, Wang W, Chen B, Fan C (2021) SA-UNet: spatial attention U-net for retinal vessel segmentation. In: 2020 25th International Conference on pattern recognition (ICPR), pp 1236–1242. https://doi.org/10.1109/ICPR48806.2021.9413346
34. Ghiasi G, Lin T-Y, Le QV (2018) DropBlock: a regularization method for convolutional networks. In: Proceedings of the 32nd International Conference on neural information processing systems (NIPS'18). Curran Associates Inc., Red Hook, NY, USA, pp 10750–10760. https://doi.org/10.5555/3327546.3327732
35. Hu J, Shen L, Sun G (2018) Squeeze-and-excitation networks. In: 2018 IEEE/CVF Conference on computer vision and pattern

recognition, pp. 7132–7141. https://doi.org/10.1109/CVPR.2018.00745
36. Woo S, Park J, Lee, J-Y, Kweon IS (2018) CBAM: convolutional block attention module. In: Ferrari V, Hebert,M, Sminchisescu, C, Weiss Y (eds) Computer Vision—ECCV 2018. ECCV 2018. Lecture Notes in Computer Science(), vol 11211. https://doi.org/10.1007/978-3-030-01234-2_1
37. Ioffe S, Szegedy C (2015) Batch normalization: Accelerating deep network training by reducing internal covariate shift. International conference on machine learning. PMLR, pp 448–456. arXiv:abs/1502.03167. https://doi.org/10.5555/3045118.3045167
38. Zhang Z, Liu Q, Wang Y (2018) Road extraction by deep residual U-Net. IEEE Geosci Remote Sens Lett 15(5):749–753. https://doi.org/10.1109/LGRS.2018.2802944
39. Cheng P, Yang Y, Yu H et al (2021) Automatic vertebrae localization and segmentation in CT with a two-stage Dense-U-Net. Sci Rep 11:22156. https://doi.org/10.1038/s41598-021-01296-1
40. Z. Wang, Z. Zhang and I. Voiculescu (2021) RAR-U-NET: a residual encoder to attention decoder by residual connections framework for spine segmentation under noisy labels. In: 2021 IEEE International Conference on Image Processing (ICIP). pp 21–25. https://doi.org/10.1109/ICIP42928.2021.9506085
41. Tran S-T, Nguyen M-H, Dang H-P, Nguyen T-T (2022) Automatic polyp segmentation using modified recurrent residual Unet network. IEEE Access 10:65951–65961. https://doi.org/10.1109/ACCESS.2022.3184773