frontiers | Frontiers in Physiology

TYPE Original Research
PUBLISHED 24 October 2022
DOI 10.3389/fphys.2022.1051808

OPEN ACCESS

EDITED BY
Zhuhuang Zhou,
Beijing University of Technology, China

REVIEWED BY
Congzhi Wang,
Shenzhen Institutes of Advanced Technology (CAS), China
Qinghua Huang,
Northwestern Polytechnical University, China

*CORRESPONDENCE
Weiwei Jiang,
jwwzjut@zjut.edu.cn

SPECIALTY SECTION
This article was submitted to Computational Physiology and Medicine, a section of the journal Frontiers in Physiology

RECEIVED 23 September 2022
ACCEPTED 10 October 2022
PUBLISHED 24 October 2022

CITATION
Jiang W, Mei F and Xie Q (2022), Novel automated spinal ultrasound segmentation approach for scoliosis visualization. *Front. Physiol.* 13:1051808. doi: 10.3389/fphys.2022.1051808

Novel automated spinal ultrasound segmentation approach for scoliosis visualization

Weiwei Jiang*, Fang Mei and Qiaolin Xie

College of Computer Science and Technology, Zhejiang University of Technology, Hangzhou, China

Scoliosis is a 3D deformity of the spine in which one or more segments of the spine curve laterally, usually with rotation of the vertebral body. Generally, having a Cobb angle (Cobb) greater than 10° can be considered scoliosis. In spine imaging, reliable and accurate identification and segmentation of bony features are crucial for scoliosis assessment, disease diagnosis, and treatment planning. Compared with commonly used X-ray detection methods, ultrasound has received extensive attention from researchers in the past years because of its lack of radiation, high real-time performance, and low price. On the basis of our previous research on spinal ultrasound imaging, this work combines artificial intelligence methods to create a new spine ultrasound image segmentation model called ultrasound global guidance block network (UGBNet), which provides a completely automatic and reliable spine segmentation and scoliosis visualization approach. Our network incorporates a global guidance block module that integrates spatial and channel attention, through which long-range feature dependencies and contextual scale information are learned. We evaluate the performance of the proposed model in semantic segmentation on spinal ultrasound datasets through extensive experiments with several classical learning segmentation methods, such as UNet. Results show that our method performs better than other approaches. Our UGBNet significantly improves segmentation precision, which can reach 74.2% on the evaluation metric of the Dice score.

KEYWORDS

ultrasound, medical image segmentation, scoliosis, 3D ultrasound image reconstruction, deep learning

1 Introduction

Scoliosis is a 3D deformity of the spine, and it includes coronal, sagittal, and axial sequence abnormalities (Konieczny et al., 2013). The commonly used detection method today is to take a standing-position, full-spine X-ray. If the frontal X-ray film shows that the spine has a lateral curvature of more than 10°, the person is diagnosed as having scoliosis (Kawchuk and Mcarthur, 1997). The causes of scoliosis include congenital, acquired, or degenerative problems, but the cause of most scoliosis

cases is unknown; this type is called idiopathic scoliosis, which is the most common type of scoliosis at present. According to statistics, about 80% of scoliosis cases belong to this category, and the difference between genders is significant, with women outnumbering men by 7:1. This condition often occurs during adolescence and is known as adolescent idiopathic scoliosis (AIS) (Weinstein et al., 2008). During the period of rapid growth, the development accelerates and then gradually deteriorates, leading to complications. For mild patients, conservative treatment with brace correction can be adopted (Negrini et al., 2011), but for severe patients, permanent spinal fusion surgery is a common method. However, permanent spinal fusion surgery greatly limits the patient's range of motion, and the complexity of the surgery is prone to complications. Therefore, the best treatment is early detection and frequent monitoring (Reamy and Slakey, 2001).

Currently, the clinical detection of scoliosis mainly relies on X-rays. However, X-rays are radioactive, and routine monitoring during rehabilitation interventions is not advisable. In addition, observing the 3D structure of the human spine on X-ray films is difficult. Magnetic resonance imaging (MRI) is a safe imaging examination based on the principle of nuclear magnetic resonance. However, MRI usually requires the patient to be in a prone position, which may cause changes in spinous process morphology (Sailer et al., 2008). Moreover, it is expensive and takes a long time, so it cannot be used on occasions with high real-time requirements. Compared with the commonly used X-ray and MRI detection methods, ultrasound has no radiation, high real-time performance, and low price (Yang et al., 2021). Ultrasound is a mechanical wave generated by mechanical vibration (Ahmed et al., 2018). The wave can enter the human body and pass through various tissues to generate echoes and form images through computer calculations. Therefore, spine ultrasound imaging has become a research hotspot in the field of spine imaging.

2D images cannot be directly used to guide the examination of scoliosis due to the limitation of visual field space in 2D images. Recent studies have indicated that 3D ultrasound has a broad application prospect in the diagnosis of scoliosis. How to visualize the 3D shape of the spine and use it for subsequent clinical research has become a hot issue. Extended field-of-view ultrasound (United States EFOV) imaging is a technique used extensively in the clinical field to attain interpretable panorama of anatomy. Huang et alproposed a novel method called double-sweep 2.5-D EFOV to better image the spinal tissues and easily compute the Cobb angle (Huang et al., 2019). Cheung et aldeveloped a 3D ultrasound system to assess AIS (Cheung et al., 2013). A novel 2.5D extended field of view method was proposed for the assessment of scoliosis (Huang et al., 2018). Zheng et aldeveloped a 3D ultrasound imaging system called Scolioscan that can be used for spine scanning (Zheng et al., 2016). In recent years, with the gradual maturity of artificial intelligence theory and applications, deep learning technology has been widely used in the field of image processing (Cai et al., 2020). Huang et alproposed a new imaging method (Huang et al., 2022) to generate the 3D structure of the human spine through tracked freehand United States scanning. Tiny-YOLOv3 (Yi et al., 2019) and K-means clustering were applied in their study to predict the spatial location of vertebral landmarks; then, they modeled the vertebrae based on the spatial position of the vertebral landmarks to form the whole spine (Huang et al., 2022). Another state-of-the-art research method was proposed by Ungi et al. (2020). They used Unet (Ronneberger et al., 2015) to segment bony features in ultrasound images and realized the visualization of 3D spine models and measurement of scoliosis degree with the help of 3DSlicer.

Inspired by these previous studies, we adopted a two-stage processing strategy to visualize a 3D model of the spine. Different from Ungi et al., we developed a novel segmentation network for spinal ultrasound images, namely, ultrasound global guidance block network (UGBNet), to achieve an accurate segmentation of 2D spinal ultrasound images. Traditional convolution neural network segmentation, such as UNet, has local receptive fields, lacks long-term dependence, and is unable to make full use of the object-to-object relationship in the global view, which may lead to potential differences between the corresponding features of pixels with the same label. At the same time, these networks do not make full use of the feature information of the intermediate convolutional layer and ignore the global context information of different scales. In the current work, the proposed network UGBNet learns long-range feature dependencies through the global guidance block (GGB) module and aggregates non-local features in a spatial-wise and channel-wise manner after processing by the GGB module to obtain accurate segmentation results. The effective information obtained from the segmentation is combined with the position information obtained from the freehand ultrasound imaging system (Cheung et al., 2015) to visualize the 3D structure of the human spine. This inexpensive approach is convenient and intuitive in displaying the spine shape, realizes the visualization of the 3D shape of the spine, and is important for doctors' follow-up diagnoses and the formulation of treatment plans.

The rest of this paper is organized as follows. Section 2 introduces the overall medical image segmentation and the detailed methods used for our spinal ultrasound image segmentation and reconstruction. The experimental settings and evaluation indicators are elaborated in Section 3. The experimental results are presented in Section 4. The discussion and conclusions are given in Sections 5 and Section 6, respectively.

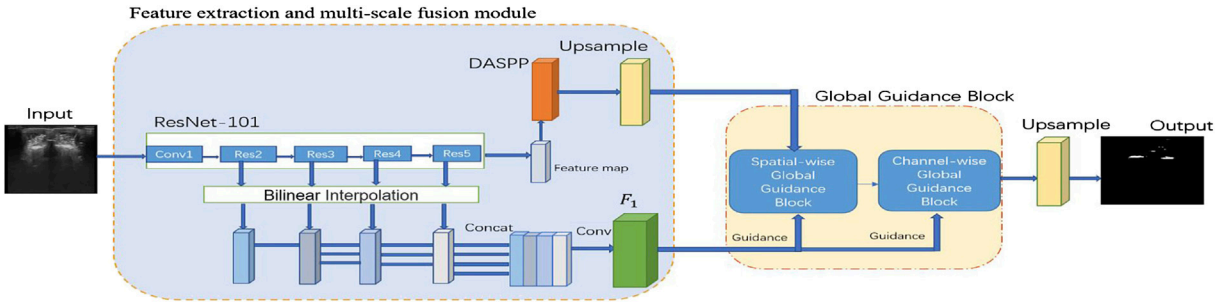

FIGURE 1
Schematic illustration of the proposed spinal ultrasound image segmentation network called UGBNet. ResNet-101 is used to extract dense local features.

2 Materials and methods

2.1 Overview

Image segmentation plays an important role in the quantitative and qualitative analyses of medical ultrasound images, and it directly affects subsequent analysis and processing (Patil and Deore, 2013). Correct segmentation guarantees the accurate extraction of diagnostic information from ultrasound images for clinical applications (Huang et al., 2021). It is also a crucial part of quantitative analysis in real-time clinical monitoring and precise positioning in computer-aided operations (Luo et al., 2021). To effectively visualize the 3D shape of the spine, we segmented the ultrasonic image. Medical ultrasound images have low image quality due to the limitation of imaging methods (Saini et al., 2010), but detailed features are an important basis for doctors' diagnoses and identification. Therefore, the details of the original ultrasound image should be preserved as much as possible even though the ultrasound image is smoothed and denoised (Huang et al., 2020). To obtain detailed features and fine segmentation results, we need to derive global features and contextual information (Chen L et al., 2017). Previous studies have suggested enlarging the receptive field by expanding convolution and pooling operations (Chen C et al., 2017) or fusing mid- and high-level features with many task-related semantic features (Ronneberger et al., 2015; Zhao et al., 2017). However, these methods cannot capture contextual information in a global view and only consider the interdependencies between spatial domains.

In this work, we developed a new network structure called UGBNet. It uses an architecture based on the ResNet network module (He et al., 2016) to integrate features and unify the feature maps generated by each ResNet building block to the same size through interpolation. Concatenate and convolution operations are performed to achieve multi-scale feature fusion and generate multi-scale feature maps. We also incorporated a GGB module (Xue et al., 2021) that integrates spatial and channel attention through which long-range feature dependencies and contextual scale information are learned. Our UGBNet can integrate deep and shallow features to generate multi-level synthetic features as the spatial and channel-wise guiding information of non-local blocks (Chen L et al., 2017) and to complement the edge details that are usually ignored by deep CNNs. Guided by multi-level comprehensive features, our UGBNet can aggregate non-local features in spatial and channel domains, effectively combine long-term non-local features provided by distant pixels in ultrasound images, and learn the semantic information of powerful non-local features for an enhanced segmentation.

Figure 1 shows the proposed UGBNet network structure. The network uses 2D spine ultrasound images as the input. First, Resnet's structural blocks are employed to extract image features and then combined with the dense atrous spatial pyramid pooling (DASPP) module (Yang et al., 2018) to expand the receptive field. Second, the GGB module is introduced to make full use of the complementary information between different CNN layers. The GGB algorithm refines features by learning long-range feature dependencies under the guidance of low-level comprehensive feature maps. The output feature map of the GGB module is used as the prediction result of our network structure. After processing by the GGB module, the spatial-wise and channel-wise non-local features are aggregated to obtain an accurate segmentation effect.

2.2 Feature extraction and multi-scale fusion module

Image segmentation can be understood as a pixel-level classification problem. To identify the category that an image belongs to, we need to distinguish it from other image categories. Feature extraction plays an important role in image recognition and classification. Traditional feature extraction methods include scale-invariant feature transform and histogram of oriented

gradient. With the development of deep learning, feature extraction through neural networks has been widely used. As one of the best approaches, ResNet has been widely adopted in image detection, segmentation, recognition, and other fields. ResNet designs a residual structure by using skip connection, which makes the network reach a deep level and endows it with an identity mapping ability and improved performance.

On the basis of the powerful feature extraction capability of ResNet, our UGBNet network structure adopts ResNet-101 (He et al., 2016) as the basic feature extraction network and uses 2D spine ultrasound images as the input. Each ResNet structure block can generate different feature maps to extract different features of the images. In particular, a DASPP module (Yang et al., 2018) is connected to the ResNet block to expand the receptive field, and its generated results are used as part of the input to the GGB module. The GGB module is discussed in Section 2.3.

To synthesize the semantic information of shallow and high-level networks, we need to carry out multi-scale feature fusion. Usually, different features can be observed at different scales to accomplish different tasks (Chen et al., 2016). With the deepening of the network layers, the receptive field of the network gradually enlarges, and the semantic expression ability is enhanced. However, this reduces the resolution of the image, and many detailed features become increasingly blurred after the convolutional operation of the multi-layer network. The convolutional neural network extracts the features of the target through layer-by-layer abstraction (Long et al., 2017). In the presence of only small local features or when the receptive field is too large, the obtained feature information is one-sided, and the possibility of obtaining too much invalid information arises. Using learned features at multiple scales helps encode global and local contexts.

In our paper, multi-scale feature prediction fusion is denoted as F_1. We adopt ResNet-101 to extract dense local features. In this situation, because the features of each scale have different resolutions, they are up-sampled to a common resolution *via* bilinear interpolation. Then, feature maps from all scales are concatenated to form a tensor, which is convolved to create multi-scale feature prediction fusion. We generate feature maps from Res-2, Res-3, Res-4, and Res-5 in ResNet and unify them to the same scale through bilinear interpolation (Mastyło, 2013), channel stacking, and convolution to form a multi-scale fusion feature map F_1, as illustrated in Figure 1. Therefore, our multi-scale fusion feature maps combine low-level details from shallow layers with high-level semantics learned in deep layers.

2.3 GGB

The traditional convolutional and recurrent operations of convolutional neural networks usually process only one local neighborhood and capture its spatial dependencies at a time (Wang et al., 2018). Although we can learn long-range dependencies by stacking convolutional layers, repeated local convolutions are time consuming. In addition, due to the limitation of imaging methods, spine ultrasound images usually contain speckles and shadows, and the signal-to-noise ratio is low (Bvsc, 2005). Diagnosis and identification pose difficulties. In this regard, we introduce GGB. The GGB module utilizes a guiding feature map to learn long-range dependencies by considering spatial and channel information, which is essential for achieving improved segmentation results.

2.3.1 Spatial-wise GGB

Figure 2 presents our spatial-wise GGB module. The output feature map of the DASPP module (Yang et al., 2018) is represented by F_X, and the guidance feature map is denoted as F_G. $L_{\alpha(x)}$, $L_{\beta(x)}$, and $L_{\gamma(x)}$ are three 1×1 convolutional layers with different parameters, and F_X is sent to them by the spatial-wise GGB module. Feature maps $\alpha_{(x)}$, $\beta_{(x)}$, and $\gamma_{(x)}$ are generated at the end. Then, matrix reshaping is performed. $\alpha_{(x)}$, $\beta_{(x)}$, and $\gamma_{(x)}$ are reshaped as $R^{hw\times c}$ matrices. In the end, we multiply the transpose of the reshaped $\alpha_{(x)}$ with the reshaped $\beta_{(x)}$ to derive a multiplication result. A softmax layer is applied to the multiplication result to calculate hw×hw spatial-wise position similarity map S_W as follows:

$$S_W = \mathrm{Softmax}\left(F_X^T L_{\alpha(x)}^T L_{\beta(x)} F_X\right).$$

The traditional sigmoid activation function is followed by a softmax layer, and it is applied to each hw × hw$F_X^T L_{\alpha(x)}^T L_{\beta(x)} F_X$. $L_{\eta(g)}$ and $L_{\rho(g)}$ are two 1×1 convolutional layers with parameters, and they are applied to guidance map F_G. Afterward, we acquire feature maps η(x)and ρ(x), reshape η(x) and ρ(x), and multiply the reshaped η(x) to the transpose of the reshaped ρ(x). Then, a softmax layer is used again, which generates another hw×hw matrix of positional similarity from guidance map F_G (denoted as M_G).

$$M_G = \mathrm{Softmax}\left(F_G^T L_{\rho(g)}^T L_{\eta(g)} F_G\right)\Big)$$

When the two similarity matrices S_W and M_G are obtained, we conduct element-wise multiplication of S_W and M_G, and a softmax layer is applied to their result. This operation generates a guided similarity matrix M_S. In the end, we multiply M_S with the feature $\gamma(x)$ to derive a new feature map F_Y, which is then added with input feature F_X to generate output feature map Y.

$$\mathrm{Y} = \gamma(x)\mathrm{Softmax}\left(S_W \cdot M_G\right) + F_X$$

2.3.2 Channel-wise GGB

When learning long-range correlations, our spatial segmentation algorithm treats each feature channel equally and ignores the correlations between different feature

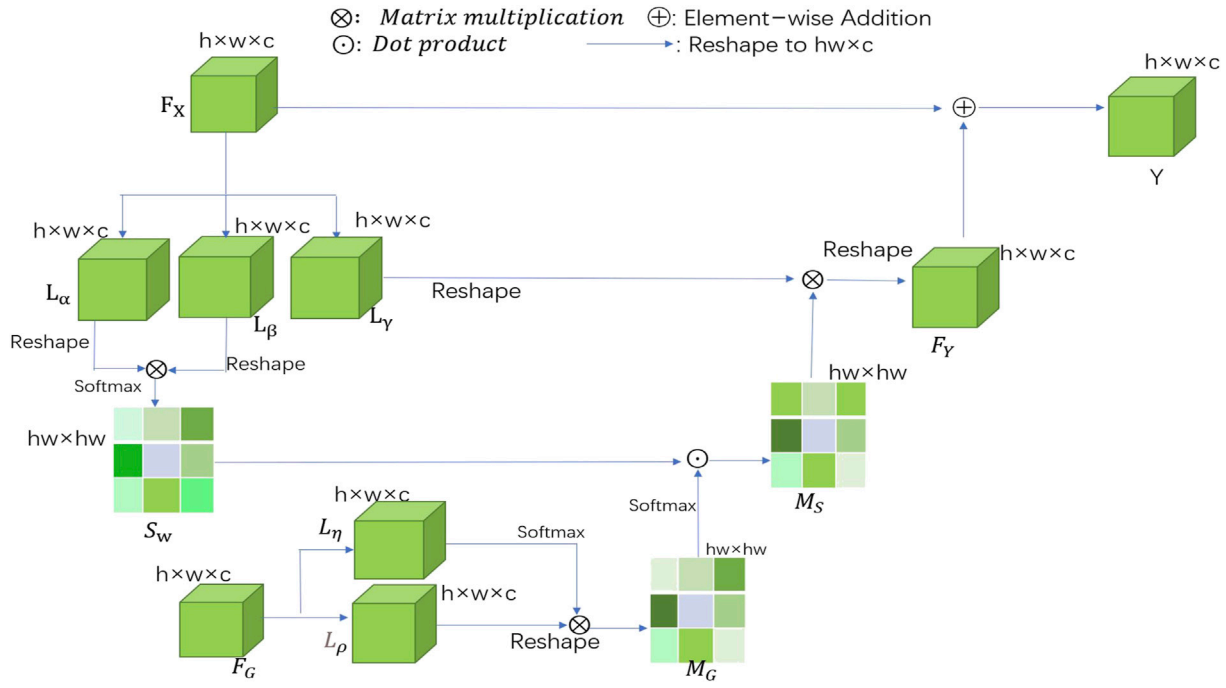

FIGURE 2
Schematic illustration of the details of spatial-wise GGB, where F_G is the guidance map and F_X is the input feature map.

channels. In recent years, many researchers have adopted strategies that allow for different contributions of different feature channels, thus achieving excellent results in many computer vision tasks (Hu et al., 2019; Lee et al., 2020). On this basis, a channel-wise GGB (channel-wise GGB) is introduced to further understand the long-range interdependencies among different feature channels. Figure 3 shows a schematic of the proposed channel-wise GGB. Feature map Y and guidance map F_G are used as two inputs to the channel-wise GGB module. Refined feature map Z is subsequently generated. In addition, feature map Y is reshaped to $R^{c\times hw}$; we multiply the reshaped Y by its transpose and use the softmax layer to obtain channel-wise similarity feature map $M_Z \in R^{c\times c}$. For input guidance feature map F_G, the informative feature channels of channel F_G are emphasized, and the less-used feature channel is suppressed using the squeeze-and-excitation block. For this purpose, we utilize global average pooling to generate channel-wise statistics λ, and the k-th layer element of the descriptor (λ) is given by

$$\lambda_k = \frac{1}{h \times w}\sum_{i=1}^{h}\sum_{j=1}^{w} F_G(i,j,k),$$

where F_G(i, j, k) represents the element of the guidance map at position (i,j,k). Two fully connected (FC) layers and a sigmoid activation function are applied to channel-wise statistics λ, thus generating coefficient vector V_c as follows:

$$V_c = \delta(P_2\phi(P_1\lambda)),$$

where P_1 and P_2 represent the parameters of the two FC layers and ϕ and δ are the ReLU and sigmoid activation functions, respectively. Next, we multiply V_c with F_G to assign different weights to the F_G channel, resulting in a refined feature map (denoted as $\hat{F}_G$). After obtaining $\hat{F}_G$, we reshape it to $R^{c\times hw}$ and multiply the reshaped $\hat{F}_G$ and the transpose of the reshaped $\hat{F}_G$. A softmax layer is then used to generate c×c similarity feature map $M_{\hat{G}}$. Subsequently, we multiply M_Z with $M_{\hat{G}}$, and a softmax layer is used for this process. At the end of this process, guided similarity map M_Q is acquired. We multiply input Y with M_Q to obtain new feature map F_Z. F_Z is added to input feature Y, which produces the output feature map Z of our channel-wise GGB.

2.4 Loss function

In this study, we use binary cross entropy (BCE) loss for network training. BCE is one of the widely used loss functions in two-class image segmentation tasks, and it reflects the direct difference between predicted masks and ground-truth labels. Its definition can be expressed as

$$\ell_{BCE} = -\sum_{(i,j)} Y(i,j)\cdot \log X(i,j) + (1 - Y(i,j))\cdot \log(1 - X(i,j)),$$

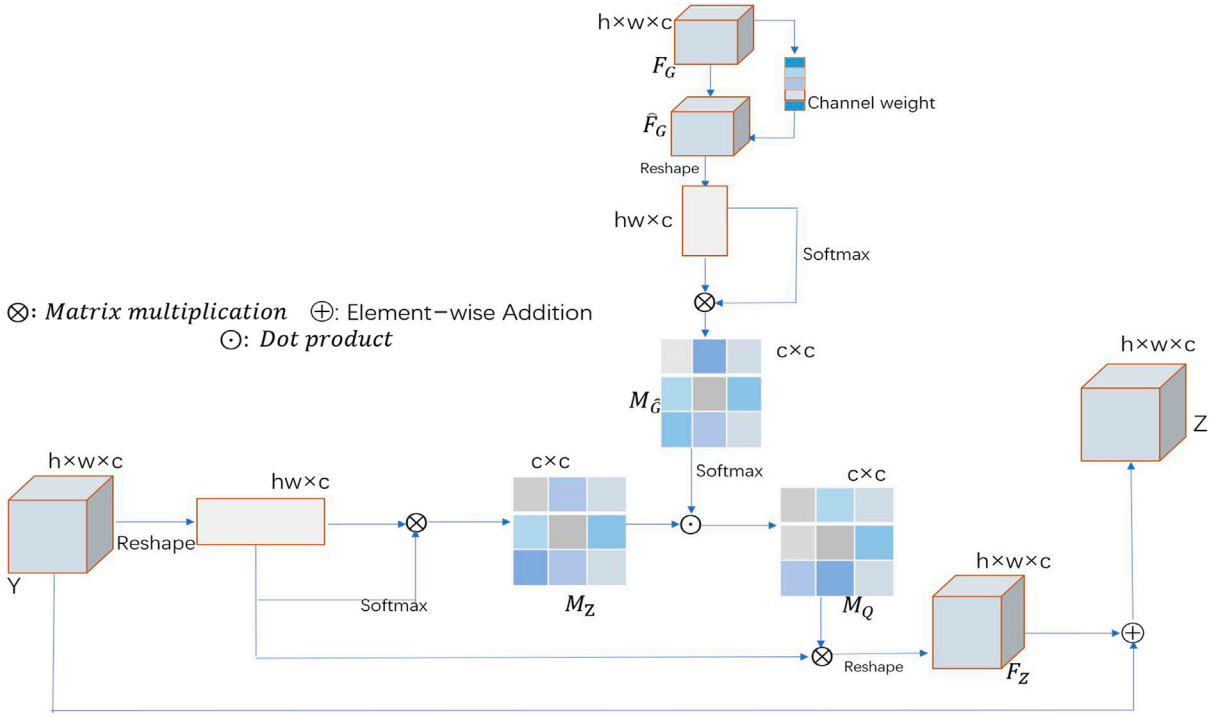

FIGURE 3
Schematic illustration of the channel-wise GGB, where F_G is the guidance map and Y is the input feature map.

where $Y(i,j) \in [0,1]$ represents the ground-truth label of pixel (i,j) and $X(i,j) \in [0,1]$ denotes the predicted masks.

3 Experiments

3.1 Experimental settings

During training, we selected the ADAM optimizer to train our network. The initial learning rate of our network was 0.001. Multiple cross-validations showed that the segmentation performance was excellent when the epoch and batch sizes were set to 50 and 8, respectively. Our experimental device was a PC with four NVIDIA Geforce RTX 2080Ti GPUs. The development environment was Ubuntu 16.04, Python 3.6, and Pytorch 1.4.0. When outputting the training results in the testing phase, we used FC CRFs (Chen et al., 2014) on the refined segmentation results outputted by the GGB module to obtain the final predicted segmentation results. FC CRFs (Chen et al., 2014) can process the classification results obtained by deep learning in consideration of the relationship between all pixels in the image. It can also optimize the rough and uncertain labels in the classified image, correct the delicate misclassification areas, and obtain detailed segmentation boundaries.

3.2 Data acquisition

Our experimental data were scanned using the freehand 3D ultrasound imaging system (Figure 4 provides an illustration). Freehand 3D ultrasound refers to a 3D ultrasound formed by using traditional 2D black-and-white ultrasound diagnostic equipment combined with a certain positioning mechanism to obtain a series of 2D ultrasound images and the corresponding spatial positions through freehand scanning and perform 3D reconstruction (Gee et al., 2003). Freehand scanning is performed with a doctor's hand-held probe, which is consistent with clinical ultrasound diagnosis and treatment applications, and the probe movement is not restricted. Images in any direction can be obtained. It is an economical, convenient, and flexible imaging method. Our study was conducted in accordance with local institutional review board standards, and all participants (or parents of participants under 18 years of age) provided written informed consent to participate in the study. A total of 102 AIS patients were recruited, and each participant could record approximately 2000 B-mode ultrasound images and their corresponding spatial data. The images we obtained were all 640×480 grayscale images, and the typical ultrasound images we used are shown in Figure 5.

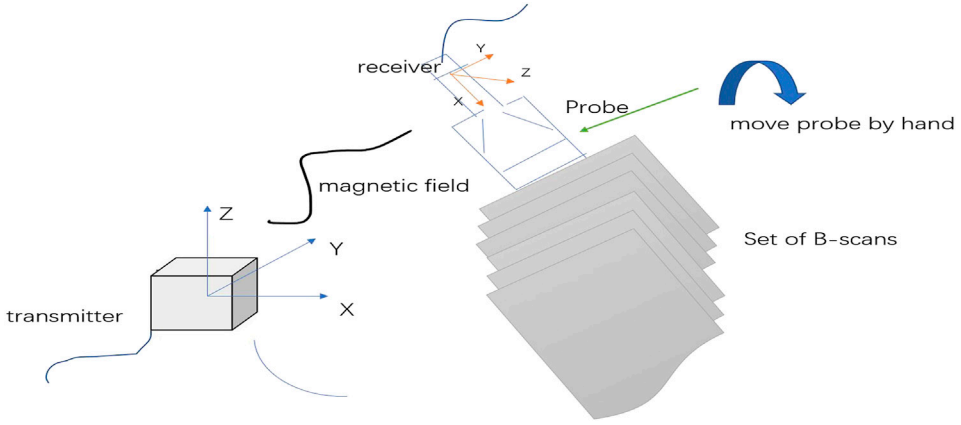

FIGURE 4
3D freehand ultrasound imaging. Freehand imaging allows the physician to move the probe freely so the B-scans can have arbitrary relative locations. The advantages of freehand scanning are low cost and scanning flexibility.

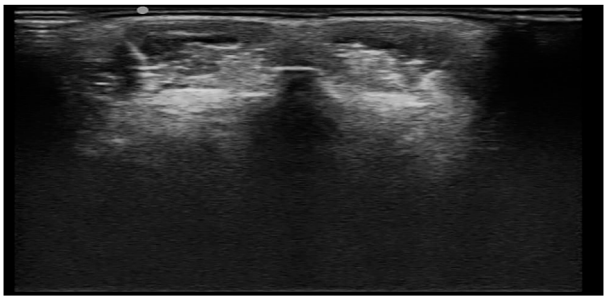

FIGURE 5
Typical spinal ultrasound image.

3.3 Evaluation metrics

The similarity between the ground truth and CNN-based segmentation results can be assessed by employing several comparison metrics. We adopted four commonly used metrics to quantitatively compare different methods of spinal ultrasound image segmentation. The four metrics were Dice coefficient (denoted as Dice from hereon), Jaccard index (denoted as Jaccard from hereon), recall, and precision. Dice and Jaccard measure the similarity between the segmentation result and the ground truth. Precision and recall compute the pixel-wise classification accuracy to evaluate the segmentation result. In general, a good segmentation result has high values of these metrics. These evaluation metrics are calculated as follows:

1 Dice coefficient:

$$\text{Dice}\ (X,Y) = \frac{2|X \cap Y|}{|X| + |Y|}$$

2 Jaccard index:

$$\text{Jaccard}\ (X,Y) = |\ \text{intersection}\ (X,Y)\ |\ /\ (\ \text{union}\ (X,Y))$$

$$= \frac{|X \bigcap Y|}{|X \bigcup Y|}$$

$$= \frac{|X \bigcap Y|}{|X| + |Y| - |X \bigcap Y|}$$

where X is the gold standard, which is the average result marked by experienced clinical experts; Y is the region segmented by the model; and X ∩ Y represents the region of overlap between the gold standard and the segmentation output of the model.

3 Recall:

$$\text{Recall} = \frac{TP}{TP + FN}$$

4 Precision:

$$\text{Precision} = \frac{TP}{TP + FP}$$

The calculation of recall and precision is associated with the true positive (TP), true negative (TN), false positive (FP), and false negative (FN) of the confusion matrix. TP is the positive image block correctly recognized by the network. FP is a negative image block that is incorrectly identified by the network as belonging to the positive image block. FN is a positive image block that is not recognized as belonging to the target image block.

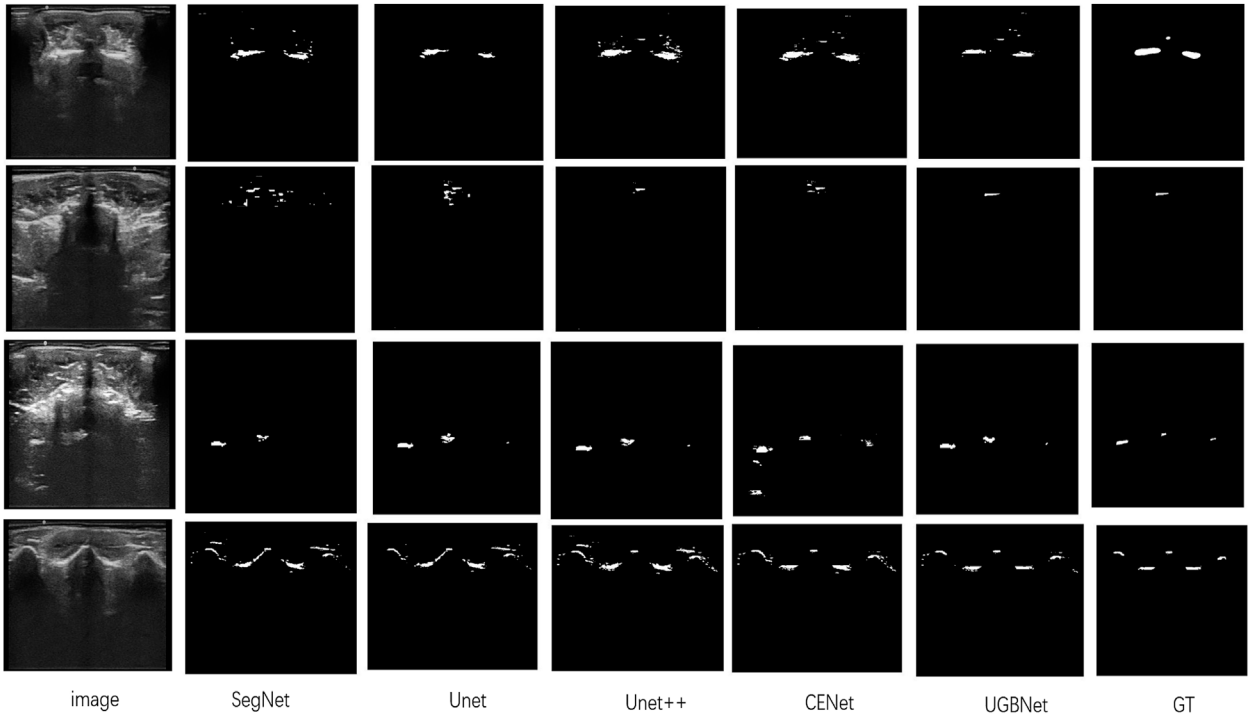

FIGURE 6
Segmentation results of different methods on our spinal ultrasound image dataset. The ground truth is denoted as GT. From left to right are our original images and the segmentation results of SegNet, Unet, Unet++, CENet, and our method. The last column is the ground truth.

4 Results

4.1 Segmentation results

To evaluate the segmentation effect of different network structures, we performed ablation experiments on our dataset. To accomplish the task of cross-validation on the dataset, we also conducted data labeling (the training dataset was labeled). We invited relevant practitioners to serve as a guide in annotating bony features, such as spinous processes, transverse processes, and ribs, in the 2D images obtained by our ultrasound scan; the features were then used as the ground truth in our experimental dataset. We compared our network against several deep-learning-based segmentation methods, including Unet (Ronneberger et al., 2015), NestedUnet (Zhou et al., 2018), SegNet (Badrinarayanan et al., 2017), and CENet (Gu et al., 2019). To ensure the fairness of the comparison, all comparative experiments were performed on the same spinal ultrasound dataset *via* four-fold cross-validation.

Visual comparison. According to the visualization results of the segmentation shown in Figure 6, our approach precisely segmented the spinous processes and laminae from the ultrasound images despite the presence of serious artifacts, whereas the other methods tended to generate over- or under-segmented results. Our network could successfully segment images with vague boundaries and detect small objects in the images. Its results were the most consistent with the ground truth among all the segmentation results.

TABLE 1 Quantitative evaluation of different methods for spinal ultrasound image segmentation.

	Dice%	Recall%	Jaccard%	Precision%
UGBNet	74.2 ± 1.2	78.5 ± 1.8	66.8 ± 1.5	79.5 ± 1.6
Unet	63.3 ± 1.6	62.2 ± 2.1	56.8 ± 1.4	68.4 ± 1.5
SegNet	61.3 ± 1.6	64.3 ± 1.8	54.3 ± 1.2	65.2 ± 1.4
Unet++	68.8 ± 1.3	71.2 ± 1.4	58.9 ± 1.6	73.5 ± 2.1
CENet	71.5 ± 1.4	75.1 ± 1.6	59.5 ± 1.7	77.2 ± 1.1

Quantitative comparison. The quantitative evaluation of the segmentation results of spinal ultrasound images produced by the different segmentation methods is presented in Table 1. Compared with the other methods, our approach achieved higher values on Dice, Jaccard, precision, and recall measurements, demonstrating the high accuracy of the proposed approach in spinal ultrasound image segmentation.

5 Discussion

At present, the clinical measurement of scoliosis is mainly based on X-rays, but the radiation of X-rays makes it difficult to be used for long-term monitoring (Kim et al., 2010). Compared with X-rays, the new spinal ultrasound imaging method is a real-time, economical, radiation-free technology (Ahmed et al., 2018). However, ultrasound images also have their inherent limitations. Given the limitations of imaging methods, ultrasound images often have acoustic artifacts, spots, and reticulated noise, which easily hide bony features, such as spinous and transverse processes, thereby making manual recognition and segmentation increasingly difficult. Inspired by Ungi et al., our research group adopted a two-stage processing strategy for the measurement and visualization of scoliosis, that is, the spine ultrasound image was segmented and recognized, the irrelevant information and noise were eliminated, and 3D visualization of spine shape was carried out. The main contribution of our study is the development of a novel segmentation network structure called UGBNet for spine ultrasound images; UGBNet performs feature extraction and multiscale fusion and incorporates a GGB module, which learns long-range feature dependencies, aggregates non-local features in spatial and channel domains, and refines the features to obtain accurate segmentation results.

Traditional spine imaging often shows the spine morphology through 3D reconstruction, which is performed directly using images obtained from ultrasound sweeps (Cheung et al., 2015). However, due to the limitations of the depth setting of the ultrasonic probe and the surrounding magnetic field, the quality of the captured 2D images cannot be guaranteed. The large amount of shadows and noise in low-quality images bring difficulties to the subsequent image recognition and 3D reconstruction visualization. This study proposed a two-stage processing strategy, that is, the bony features in the spine ultrasound image are recognized and segmented, followed by 3D reconstruction and visualization of the spine. This two-stage processing strategy can minimize the interference of irrelevant information, such as acoustic artifacts and speckle noise, and has positive significance for the subsequent visualization of 3D spine morphology and measurement of scoliosis degree. In the segmentation and recognition of bony features of spine ultrasound images, we improved the segmentation algorithm and proposed the UGBNet network structure. Multiple qualitative and quantitative experiments showed that our method achieved higher values of Dice, precision, and other evaluation metrics compared with traditional image segmentation algorithms, such as UNet.

However, our method cannot guarantee accurate segmentation of all spinal ultrasound images. In terms of ultrasonic data acquisition, due to the inexperience of the operator, some scanning problems may arise, resulting in the poor quality of the collected 2D ultrasonic images and the presence of abundant shadows and noise (Rohling et al., 1999). Our methods are often inadequate when dealing with such images. In our future research, we will consider preprocessing the acquired 2D ultrasound image to enhance the weight of the structure of interest, find ways to improve the image contrast and image quality, and lay a good foundation for subsequent research. In addition, the performance of deep learning networks needs to be tested in a larger and more diverse patient population than the current one (Cai et al., 2020). Our sample size is relatively small, and continued large-scale clinical trials are needed to validate the feasibility of using the proposed method in the diagnosis, treatment, and screening of scoliosis.

6 Conclusion

In summary, we propose a novel spinal ultrasound image segmentation network called UGBNet, which can accurately segment and identify bony features, such as spinous and transverse processes, in spinal ultrasound images. The proposed network considers long-range dependencies in a spatial-wise and channel-wise manner and embeds contextual information from different layers. Our method can be used as the first step in a two-stage processing strategy for spinal ultrasound 3D imaging and scoliosis measurement, which is important for subsequent visualization of spinal 3D morphology and scoliosis measurement. Our approach is radiation-free and inexpensive, and it provides a new idea for the clinical measurement and treatment of scoliosis. It is a feasible alternative to current approaches that use X-ray as the main diagnostic method, and we look forward to its large-scale promotion in the future.

Data availability statement

The raw data supporting the conclusion of this article will be made available by the authors, without undue reservation.

Ethics statement

The studies involving human participants were reviewed and approved by Health Subjects Ethics Subcommittee at PolyU. The patients/participants provided their written informed consent to participate in this study.

Author contributions

WJ developed the research question and conceived this study. FM conducted the relevant experiment and data analysis and drafted the manuscript. QX collected data and recruited volunteers.

Funding

This work was supported in part by the Natural Science Foundation of Zhejiang Province (LY20H180006) and the National Natural Science Foundation of China (61701442).

Acknowledgments

The authors would like to thank the participants for their dedication, without which this study would not have been possible. We are grateful to the volunteers who volunteered to participate in this study for their outstanding contributions to the data collection.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

References

Ahmed, A. S., Ramakrishnan, R., Ramachandran, V., Ramachandran, S. S., Phan, K., and Antonsen, E. L. (2018). Ultrasound diagnosis and therapeutic intervention in the spine. *J. Spine Surg.* 4 (2), 423–432. doi:10.21037/jss.2018.04.06

Badrinarayanan, V., Kendall, A., and Cipolla, R. (2017). SegNet: A deep convolutional encoder-decoder architecture for image segmentation. *IEEE Trans. Pattern Anal. Mach. Intell.* 39, 2481–2495. doi:10.1109/TPAMI.2016.2644615

Bvsc, M. (2005). Kirberger R. Imaging artifacts in diagnostic ultrasound—a review. *Veterinary Radiology Ultrasound* 36 (4), 297–306. doi:10.1111/j.1740-8261.1995.tb00266.x

Cai, L., Gao, J., and Zhao, D. (2020). A review of the application of deep learning in medical image classification and segmentation. *Ann. Transl. Med.* 8 (11), 713. doi:10.21037/atm.2020.02.44

Chen C, L.-C., Papandreou, G., Kokkinos, I., Murphy, K., and Yuille, A. L. (2017). Deeplab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected crfs. *IEEE Trans. Pattern Anal. Mach. Intell.* 40 (4), 834–848. doi:10.1109/TPAMI.2017.2699184

Chen L, L., Zhang, H., Xiao, J., Nie, L., Shao, J., Liu, W., et al. (2017). "Sca-cnn: Spatial and channel-wise attention in convolutional networks for image captioning," in Proceedings of the IEEE conference on computer vision and pattern recognition), Honolulu, HI, July 21–26, 2017. 5659–5667.

Chen, L.-C., Papandreou, G., Kokkinos, I., Murphy, K., and Yuille, A. L. (2014). Semantic image segmentation with deep convolutional nets and fully connected crfs. *arXiv preprint arXiv:1412.7062*.

Chen, L. C., Yang, Y., Wang, J., Xu, W., and Yuille, A. L. (2016). "Attention to scale: Scale-aware semantic image segmentation," in Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, June 27–30, 2016. 3640–3649.

Cheung, C.-W. J., Law, S.-Y., and Zheng, Y.-P. (2013). "Development of 3-D ultrasound system for assessment of adolescent idiopathic scoliosis (AIS): And system validation," in Proceedings of the 2013 35th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC): IEEE), Osaka, Japan, 03-07 July 2013, 6474–6477.

Cheung, C.-W. J., Zhou, G.-Q., Law, S.-Y., Lai, K.-L., Jiang, W.-W., and Zheng, Y.-P. (2015). Freehand three-dimensional ultrasound system for assessment of scoliosis. *J. Orthop. Transl.* 3 (3), 123–133. doi:10.1016/j.jot.2015.06.001

Gee, A., Prager, R., Treece, G., and Berman, L. (2003). Engineering a freehand 3D ultrasound system. *Pattern Recognit. Lett.* 24 (4-5), 757–777. doi:10.1016/s0167-8655(02)00180-0

Gu, Z., Cheng, J., Fu, H., Zhou, K., Hao, H., Zhao, Y., et al. (2019). CE-Net: Context encoder network for 2D medical image segmentation. *IEEE Trans. Med. Imaging* 38 (10), 2281–2292. doi:10.1109/tmi.2019.2903562

He, K., Zhang, X., Ren, S., and Sun, J. (2016). "Deep residual learning for image recognition," in Proceedings of the IEEE conference on computer vision and pattern recognition, Las Vegas, NV, June 27–30, 2015, 770–778.

Hu, Y., Guo, Y., Wang, Y., Yu, J., Li, J., Zhou, S., et al. (2019). Automatic tumor segmentation in breast ultrasound images using a dilated fully convolutional network combined with an active contour model. *Med. Phys.* 46 (1), 215–228. doi:10.1002/mp.13268

Huang, Q., Deng, Q., Li, L., Yang, J., and Li, X. (2019). Scoliotic imaging with a novel double-sweep 2.5-dimensional extended field-of-view ultrasound. *IEEE Trans. Ultrason. Ferroelectr. Freq. Control* 66 (8), 1304–1315. doi:10.1109/TUFFC.2019.2920422

Huang, Q., Huang, Y., Luo, Y., Yuan, F., and Li, X. (2020). Segmentation of breast ultrasound image with semantic classification of superpixels. *Med. Image Anal.* 61, 101657. doi:10.1016/j.media.2020.101657

Huang, Q., Luo, H., Yang, C., Li, J., Deng, Q., Liu, P., et al. (2022). Anatomical prior based vertebra modelling for reappearance of human spines. *Neurocomputing* 500, 750–760. doi:10.1016/j.neucom.2022.05.033

Huang, Q., Miao, Z., Zhou, S., Chang, C., and Li, X. (2021). Dense prediction and local fusion of superpixels: A framework for breast anatomy segmentation in ultrasound image with scarce data. *IEEE Trans. Instrum. Meas.* 70 (70-), 1–8. doi:10.1109/tim.2021.3088421

Huang, Q., Zeng, Z., and Li, X. (2018). 2.5-D extended field-of-view ultrasound. *IEEE Trans. Med. Imaging* 37, 851–859. doi:10.1109/tmi.2017.2776971

Kawchuk, G., and Mcarthur, R. (1997). Scoliosis quantification: An overview. *Jcca.journal Can. Chiropr. Association.journal De Lassociation Chiropratique Can.* 41 (3). 137-144.

Kim, H., Kim, H. S., Moon, E. S., Yoon, C. S., Chung, T. S., Song, H. T., et al. (2010). Scoliosis imaging: What radiologists should know. *Radiographics* 30 (7), 1823–1842. doi:10.1148/rg.307105061

Konieczny, M. R., Senyurt, H., and Krauspe, R. (2013). Epidemiology of adolescent idiopathic scoliosis. *J. Child. Orthop.* 7 (1), 3–9. doi:10.1007/s11832-012-0457-4

Lee, H., Park, J., and Hwang, J. Y. (2020). Channel attention module with multiscale grid average pooling for breast cancer segmentation in an ultrasound image. *IEEE Trans. Ultrason. Ferroelectr. Freq. Control* 67 (7), 1344–1353. doi:10.1109/TUFFC.2020.2972573

Long, J., Shelhamer, E., and Darrell, T. (2017). Fully convolutional networks for semantic segmentation. *IEEE Trans. Pattern Anal. Mach. Intell.* 39, 640–651. doi:10.1109/TPAMI.2016.2572683

Luo, Y., Huang, Q., and Li, X. (2021). Segmentation information with attention integration for classification of breast tumor in ultrasound image. *Pattern Recognit.* 124 (1), 108427. doi:10.1016/j.patcog.2021.108427

Mastyło, M. (2013). Bilinear interpolation theorems and applications. *J. Funct. Analysis* 265 (2), 185–207. doi:10.1016/j.jfa.2013.05.001

Negrini, S., Minozzi, S., Bettany-Saltikov, J., Zaina, F., Chockalingam, N., Grivas, T. B., et al. (2011). Braces for idiopathic scoliosis in adolescents. *Spine* 5 (4), 1681–1720. doi:10.1097/BRS.0b013e3181dc48f4

Patil, D. D., and Deore, S. G. (2013). Medical image segmentation: A review. *Int. J. Comput. Sci. Mob. Comput.* 2 (1), 22–27.

Reamy, B. V., and Slakey, J. B. (2001). Adolescent idiopathic scoliosis: Review and current concepts. *Am. Fam. Physician* 64 (1), 111–116.

Rohling, R., Gee, A., and Berman, L. (1999). A comparison of freehand three-dimensional ultrasound reconstruction techniques. *Med. Image Anal.* 3 (4), 339–359. doi:10.1016/s1361-8415(99)80028-0

Ronneberger, O., Fischer, P., and Brox, T. (2015). "U-net: Convolutional networks for biomedical image segmentation," in *International Conference on Medical image computing and computer-assisted intervention* (Berlin, Germany: Springer), 234–241.

Sailer, M., Fazekas, F., Gass, A., Kappos, L., Radue, E. W., Rieckmann, P., et al. (2008). Cerebral and spinal MRI examination in patients with clinically isolated syndrome and definite multiple sclerosis. *Rofo* 180 (11), 994–1001. doi:10.1055/s-2008-1027817

Saini, K., Dewal, M., and Rohit, M. (2010). Ultrasound imaging and image segmentation in the area of ultrasound: A review. *Int. J. Adv. Sci. Technol.* 24, 41–60.

Ungi, T., Greer, H., Sunderland, K. R., Wu, V., Baum, Z. M., Schlenger, C., et al. (2020). Automatic spine ultrasound segmentation for scoliosis visualization and measurement. *IEEE Trans. Biomed. Eng.* 67 (11), 3234–3241. doi:10.1109/TBME.2020.2980540

Wang, X., Girshick, R., Gupta, A., and He, K. (2018). "Non-local neural networks," in Proceedings of the IEEE conference on computer vision and pattern recognition, Salt Lake City, UT, USA, 18-23 June 2018, 7794–7803.

Weinstein, S. L., Dolan, L. A., Cheng, J. C., Danielsson, A., and Morcuende, J. A. (2008). Adolescent idiopathic scoliosis. *Lancet* 371 (9623), 1527–1537. doi:10.1016/s0140-6736(08)60658-3

Xue, C., Zhu, L., Fu, H., Hu, X., Li, X., Zhang, H., et al. (2021). Global guidance network for breast lesion segmentation in ultrasound images. *Med. Image Anal.* 70, 101989. doi:10.1016/j.media.2021.101989

Yang, C., Jiang, M., Chen, M., Fu, M., and Huang, Q. (2021). Automatic 3-D imaging and measurement of human spines with a robotic ultrasound system. *IEEE Trans. Instrum. Meas.* 70 (99), 1–13. doi:10.1109/tim.2021.3085110

Yang, M., Yu, K., Zhang, C., Li, Z., and Yang, K. (2018). "Denseaspp for semantic segmentation in street scenes," in Proceedings of the IEEE conference on computer vision and pattern recognition), Salt Lake City, UT, USA, 18-23 June 2018, 3684–3692.

Yi, Z., Yongliang, S., and Jun, Z. (2019). An improved tiny-yolov3 pedestrian detection algorithm. *Optik* 183, 17–23. doi:10.1016/j.ijleo.2019.02.038

Zhao, H., Shi, J., Qi, X., Wang, X., and Jia, J. (2017). "Pyramid scene parsing network," in Proceedings of the IEEE conference on computer vision and pattern recognition), Honolulu, HI, USA, 21-26 July 2017, 2881–2890.

Zheng, Y. P., Lee, T. Y., Lai, K. L., Yip, H. K., Zhou, G. Q., Jiang, W. W., et al. (2016). A reliability and validity study for scolioscan: A radiation-free scoliosis assessment system using 3D ultrasound imaging. *Scoliosis Spinal Disord.* 11 (1), 13. doi:10.1186/s13013-016-0074-y

Zhou, Z., Rahman Siddiquee, M. M., Tajbakhsh, N., and Liang, J. (2018). "Unet++: A nested u-net architecture for medical image segmentation," in *Deep learning in medical image analysis and multimodal learning for clinical decision support* (Berlin, Germany: Springer), 3–11.