In-depth learning of automatic segmentation of shoulder joint magnetic resonance images based on convolutional neural networks

Xinhong Mu , Yi Cui , Rongpeng Bian , Long Long , Daliang Zhang , Huawen Wang , Yidong Shen , Jingjing Wu , Guoyou Zou

PII: S0169-2607(21)00399-0
DOI: https://doi.org/10.1016/j.cmpb.2021.106325
Reference: COMM 106325

To appear in: *Computer Methods and Programs in Biomedicine*

Received date: 28 December 2020
Accepted date: 25 July 2021

Please cite this article as: Xinhong Mu , Yi Cui , Rongpeng Bian , Long Long , Daliang Zhang , Huawen Wang , Yidong Shen , Jingjing Wu , Guoyou Zou , In-depth learning of automatic segmentation of shoulder joint magnetic resonance images based on convolutional neural networks, *Computer Methods and Programs in Biomedicine* (2021), doi: https://doi.org/10.1016/j.cmpb.2021.106325

Highlights

- CNN is used to segment multiple bone joints to assist medical diagnosis.
- Image segmentation using local perception and features.
- A loss function is added to adjust the position of the candidate area.
- Fine segmentation of the edges of the bone areais is performed to prevent confusion with fat.

In-depth learning of automatic segmentation of shoulder joint magnetic resonance images based on convolutional neural networks

Xinhong Mu[1,2], Yi Cui[1,2], Rongpeng Bian[1,2], Long Long[1,2], Daliang Zhang[1,2], Huawen Wang[1,2], Yidong Shen[1,2], Jingjing Wu[1,2], Guoyou Zou [1,2,*]

1.Yancheng First Hospital, Affiliated Hospital of Nanjing University Medical School

2.The First People's Hospital of Yancheng

* Corresponding author: Guoyou Zou

Abstract

Objective: Magnetic resonance imaging (MRI) is gradually replacing computed tomography (CT) in the examination of bones and joints. The accurate and automatic segmentation of the bone structure in the MRI of the shoulder joint is essential for the measurement and diagnosis of bone injuries and diseases. The existing bone segmentation algorithms cannot achieve automatic segmentation without any prior knowledge, and their versatility and accuracy are relatively low. For this reason, an automatic segmentation algorithm based on the combination of image blocks and convolutional neural networks is proposed.

Methods: First, we establish 4 segmentation models, including 3 U-Net-based bone segmentation models (humeral segmentation model, joint bone segmentation model, humeral head and articular bone segmentation model as a whole) and a block-based Alex Net segmentation model; Then we use 4 segmentation models to obtain the candidate bone area, and accurately detect the location area of the humerus and joint bone by voting. Finally, the Alex Net segmentation model is further used in the detected bone area to segment the bone edge with the accuracy of the pixel level.

Results: The experimental data is obtained from 8 groups of patients in the orthopedics department of our hospital. Each group of scan sequence includes about 100 images, which have been segmented and labeled. Five groups of patients were used for training and five-fold cross-validation, and three groups of patients were used to test the actual segmentation effect. The average accuracy of Dice Coefficient, Positive Predicted Value (PPV) and Sensitivity reached 0.91 ± 0.02, respectively. 0.95 ± 0.03 and 0.95 ± 0.02.

Conclusions: The method in this paper is for a small sample of patient data sets, and only through

deep learning on 2D medical images, very accurate shoulder joint segmentation results can be obtained, provide clinical diagnostic guidance to orthopedics. At the same time, the proposed algorithm framework has a certain versatility and is suitable for the precise segmentation of specific organs and tissues in MRI based on a small sample data.

Keywords: Deep learning; Medical image segmentation; Convolutional neural network; Magnetic resonance imaging; Orthopedic diagnosis

1 Introduction

Computed tomography (CT) uses computer technology and cross-sectional projection methods to penetrate X-rays through the tissues of each axial layer of the human body. It has high density resolution and is 100 times stronger than ordinary X-rays.[1] Therefore, CT has certain advantages for bone imaging. It is more sensitive and clear than other imaging methods. It is very easy to use computer technology to accurately obtain bone edges for segmentation, evaluation and diagnosis. However, the amount of X-ray exposure to a CT examination is much greater than that of an X-ray examination, which is more harmful to the human body[2]. The exposure caused by the CT examination may also increase the risk of cancer. On the other hand, magnetic resonance imaging (MRI) uses the principle of nuclear magnetic resonance, which is safer to use and has higher soft tissue resolution. With the further development of MRI technology and the strengthening of patients' awareness of self-care, its application has become more and more extensive. In fact, more and more bone detection and diagnosis are beginning to use MRI[3].

In recent years, due to the increase in the use of computers and mobile phones and the increase in use time, the incidence of shoulder joint diseases has continued to increase. There is an urgent need for computer-assisted automatic segmentation to accelerate the diagnosis and treatment process. For general traditional segmentation algorithms, such as region growth or level set, etc., segmentation parameters cannot be adaptively adjusted automatically according to different types of images or different regions of the same type of image. Therefore, a more common approach is to use preliminary segmentation such as region growth or level set in multiple local areas, and then manually adjust to achieve the effect of semi-automatic segmentation[4].

MRI is a type of tomography, which uses magnetic resonance phenomena to obtain electromagnetic signals from the human body and reconstruct human body information. MRI technology has some things in common with other tomography technologies, for example, they can show the distribution of a certain physical quantity (such as density) in space. At the same time, it also has its own characteristics. MRI can obtain tomographic images in any direction, three-dimensional volume images, and even four-dimensional images of space-spectrum distribution.

In MRI, the internal area of the bone and the air, fat, and some soft tissues all present a similar gray-black color, plus a lower image signal-to-noise ratio and partial volume effect. Therefore, it is difficult to automatically and accurately segment the clinically valuable humerus and joint bone in the shoulder joint[5].

Similar to PET and SPECT imaging the magnetic resonance signal used for imaging comes directly from the object itself. It can also be said that MRI is also a type of emission tomography. But unlike PET and SPECT, MRI can be imaged without injecting radioisotopes. This also makes MRI technology safer.

From the magnetic resonance image, we can get a variety of physical properties of matter, such as proton density, spin-lattice relaxation time T1, spin-spin relaxation time T2, diffusion coefficient,

magnetic susceptibility, chemical shift, etc. Compared with other imaging technologies, MRI methods are more diverse, imaging principles are more complicated, and the information obtained is more abundant. Therefore, MRI has become a popular research direction in medical imaging.

The main advantages of MRI image segmentation are that the soft tissue imaging effect is superior; the spatial resolution is high; it is harmless to the human body; the scanning angle is flexible; there is no bone artifacts, which allows it to be a valuable imaging modality for orthopedic research.

2 Methodology

2.1 Overview of the segmentation system

Convolutional Neural Network (CNN) is an important content in the field of artificial intelligence. In recent years, it has been successfully applied in various industries and has achieved a lot of research results in machine vision such as image detection, positioning, classification and segmentation. The convolutional neural network is based on the traditional neural network, which considers the spatial position of the pixel, plus the degree of interest (weight) set manually, to achieve weight sharing. In this way, you can train a deeper network structure, extract more abstract image features, greatly reduce the number of neuron parameters, and get better results and higher efficiency.

This paper proposes a method combining block-based and convolutional neural networks to accurately and automatically segment the humerus and joint bones in the MRI of the shoulder joint. The basic process of the entire system is shown in Figure 1 and Figure 2. Figure 1 is a schematic diagram of the generation of the segmentation model, and Figure 2 is a flowchart of the use of the segmentation model for bone region detection and edge fine segmentation. The final 3D visualization in Figure 2 is provided for measurement and diagnosis. The parts marked in red and blue are bones.

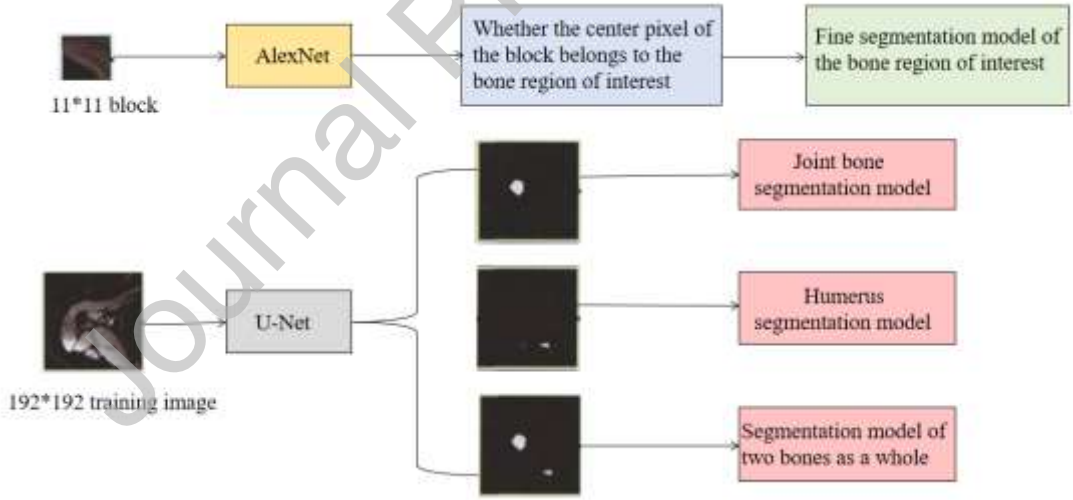


Figure 1 Generated segmentation model frame diagram

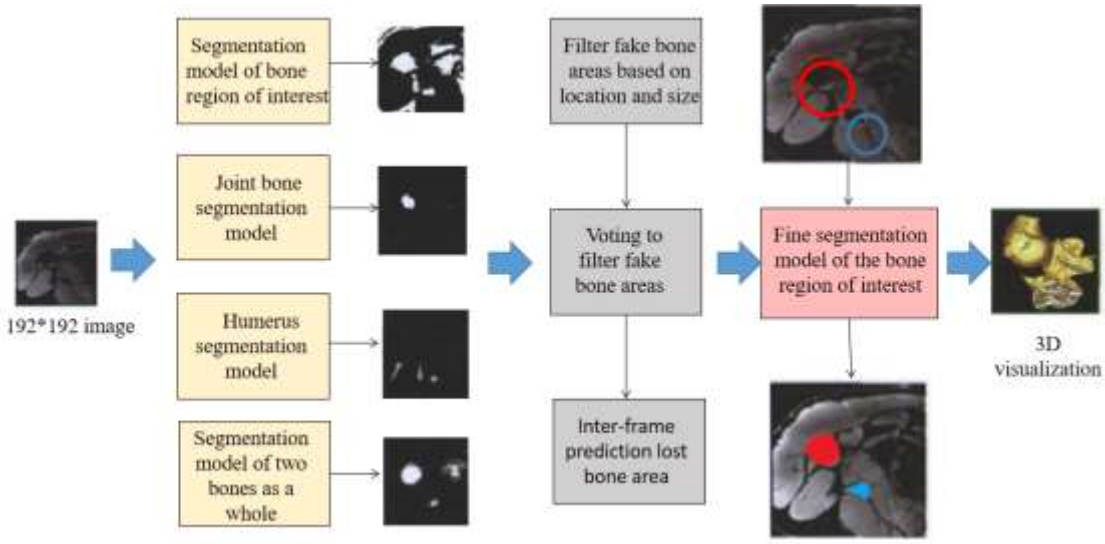


Figure 2 Frame diagram of segmentation process

2.2 Convolution operation

The conventional neural network is shown in Figure 3. We input a single vector to the neural network and transform it through a series of hidden layers made up of neurons. Each hidden layer is fully connected to all neurons in the previous layer of the network, and the neurons in a single layer work independently and do not share connections. The last layer is called the output layer. In the classification task, it represents the score of each category.

Convolutional neural network uses the conv layer and pool layer. The layers are locally connected. The conv layer and pool layer are composed of learnable weight w and bias value b neurons. Each neuron receives input data and performs dot product operations.The entire network calculation process can still be expressed as a differentiable function: from the original image at the input to the classification score at the output. As shown in Figure 3, input image data, and the convolution layer performs convolution processing.The neurons of the convolutional neural network are arranged in three dimensions (width, height, depth).

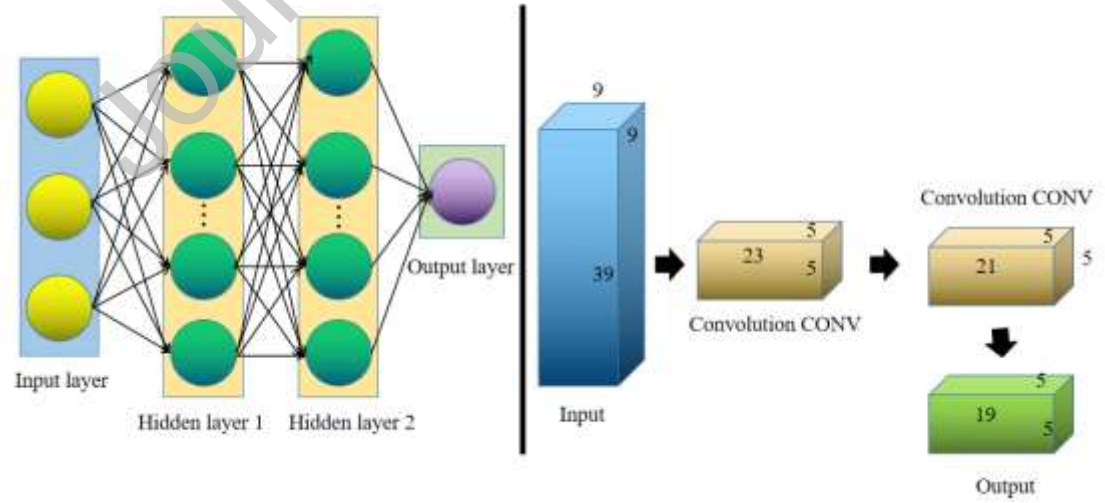


Figure 3 Schematic diagram of conventional three-layer neural network

The core module of the convolutional neural network is the conv layer, which undertakes most of the computing tasks. The parameters of the conv layer are composed of a set of learnable filters filter.

Each filter (convolution kernel) is small in space (along the width and height directions), but can be extended to the full depth of the input data. Each layer is obtained by convolution operation of the filter and the output result of the previous layer. We use filters of the same size to perform convolution calculation with the local perception area with a specific step size to extract similar features. Different size filters can extract different features. As shown in Figure 4, the convolution calculation process of a two-dimensional single-channel image is shown. The filter used in the figure is 3×3, the image size is 4×4, and the step size is 1. In image processing, the main function of convolution operation is to enhance the original signal, reduce noise, edge detection and feature extraction.

The pool layer generally connected to the conv layer, can simplify the output of the convolutional layer, compress the feature map, and extract the main features of the feature map. The pooling layer greatly reduces the number of parameters in the neural network and reduces the amount of calculation. Pooling operation refers to replacing the value of a certain area (filter range) in the image with only one pixel value, including two pooling methods of maximum pooling and average pooling.As shown in Figure 5, on the 4×4 feature map, a 2×2 filter is used with a step size of 2, and the maximum pooling operation is performed. The introduction of pooling operation in the neural network allows the model to greatly reduce the parameters while improving the generalization ability of the network, and it also has the image translation and rotation invariance.

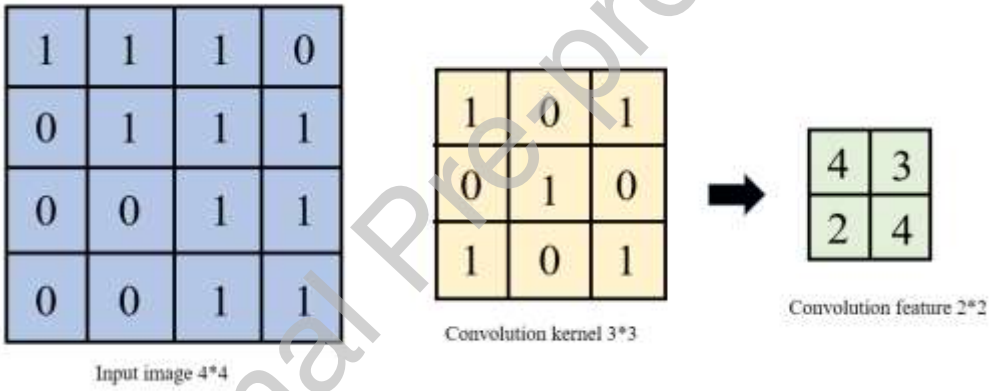


Figure 4 Schematic diagram of two-dimensional convolution

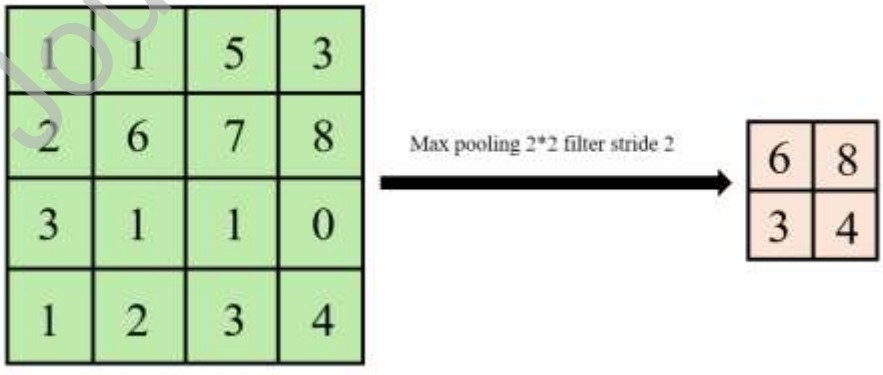


Figure 5 Two-dimensional pooling operation diagram

2.3 Bone area detection

2.3.1 Generate candidate bone regions

According to the analysis of the above characteristics, it can be concluded that a single segmentation model is difficult to accurately segment the articular bone and humeral head region at one

time. The segmentation results of all models are used as candidate bone regions, and the rectangular frame range of all candidate regions is obtained. The specific steps are as follows: (1) The segmentation output of the AlexNet model is directly a binary image, 255 represents the bone area, and 0 represents the non-bone area. (2) The output of U-Net model segmentation is the probability value of each pixel belonging to the corresponding bone area, which is multiplied by 255 and converted into a grayscale image, and converted into a binary image through adaptive binarization. The same 255 represents the bone Area, 0 means non-bone area. (3) Perform morphological operations on all segmented binary images: first erode twice, then expand twice, with a core size of 3 × 3 pixels. (4) Obtain the smallest rectangle of the contours of all segmented regions as the candidate bone region.

2.3.2 Obtain effective bone area

In the candidate bone area, first we filter according to the position to remove the rectangular frame at the boundary position, and then we filter again according to the size to remove the rectangular frame that is too large or too small. In the filtering results, the voting mechanism is adopted. If 60% of the area of a candidate area exists in the segmentation results of at least 3 models, all the candidate areas where it is located are merged, that is, the smallest outer frame of the rectangular frame of these candidate areas is obtained that is then perceived as the final effective bone area.

In the above process of obtaining effective bone regions, very few bone regions only exist in one or two prediction model segmentation results, so they will be voted in the voting process. Due to the continuity of MRI scanning in the time direction, the effective bone area has a very small change in a short time, so the lost bone area can be restored by the method of inter-frame prediction.

If there is no bone segmentation area in the current image, just retrieve the next frame and the previous frame respectively for the first frame and the last frame, and obtain the two frames before and after the others. If there is an empty area image in the previous frame, then move forward or backward by one frame. We predict the bone area in the current image by the average value of the bone rectangular area in the front and back prediction frames. If there is only one bone object in the current image, but there are two bone objects in the front and rear frames, the rectangular area where there is no object is still predicted by the average between frames.

2.4 Overview of different image segmentation methods

The U-net network is a CNN-based image segmentation network, mainly used for medical image segmentation. When the network was first proposed for cell wall segmentation, it has been excellent in lung nodule detection and blood vessel extraction on the retina at the fundus. Performance.The original U-net network structure mainly consists of a convolutional layer, a maximum pooling layer (downsampling), a deconvolution layer (upsampling), and the ReLU nonlinear activation function.U-net's network structure does not involve any fully connected layers. At the same time, the results of downsampling are used in the upsampling process, so that there can be shallow simple features in the deep convolution, making the convolution input more naturally obtained results are more able to reflect the original information of the image.

AlexNet uses CNN in a deeper and wider network, and its effect classification accuracy is higher than that of the previous LeNet. It uses ReLU instead of Sigmoid, which can train faster, and at the same time solves the problem of sigmoid's gradient disappearance in a deeper network, or gradient dispersion.It can randomly ignore some neurons in the process of image segmentation to avoid overfitting.In the previous CNNs, the average pooling layer was commonly used. AlexNet all uses the maximum pooling layer, which avoids the blurring effect of the average pooling layer, and the step size is smaller than the size of the pooling core, so the output of the pooling layer. There is overlap between

them, which enhances the richness of features.AlexNet proposed the LRN layer, which can achieve local response normalization, and create a competition mechanism for local neurons, which makes the value of the response smaller and suppresses the smaller feedback. AlexNet uses the GPU to accelerate the neural network. The training ensures the accuracy of image segmentation.

2.5 Fine edge segmentation

In order to cover all the bone objects of interest without introducing too much noise in the surrounding area, the edge of the generated effective bone area needs to be expanded to a certain extent, here is an expansion of 10 pixels. Using the predicted binary mask image of the AlexNet segmentation model again, we find the longest closed contour in the extended bone rectangle area of the binary mask image as the precise edge of the bone.

2.6 Loss function

Equation (1) is the loss function of the entire network architecture, which is mainly divided into two parts. The first part is the target classification loss, and the second part is the frame regression loss, which is used to adjust the position of the candidate area.

$$L(\{p_i\},\{t_i\})=\frac{1}{N_{cls}}\sum_i L_{cls}(p_i,p_i^*)+\lambda\frac{1}{N_{reg}}\sum_i p_i^* L_{reg}(t_i,t_i^*) \quad (1)$$

In the Equation (1), i is the index of the anchor boxes in each mini-batch in the network training, and pi is the probability of predicting whether the anchor boxes are positive samples (foreground) or negative samples (background); p_i^* is the real label corresponding to the anchor boxes. If the anchor boxes are positive samples, p_i^* is 1, and if the anchor boxes are negative samples, then p_i^* is 0; $t_i=\{t_x,t_y,t_w,t_h\}$ represents A four-dimensional vector representing the predicted offset of the anchor boxes with index i. The dimension of t_i^* is consistent with the ti dimension, and it also represents a vector, which represents the offset of the positive sample anchor boxes with index i from the ground truth bound; N_{cls} represents the total number of anchor boxes obtained at the end of the entire cls branch, which is 256 in this experiment; $L_{cls}(p_i, p_i^*)$ represents the log loss function of the two classifications (log loss); N_{reg} is the total number of anchor boxes obtained at the end of the reg branch, which is 256 in this experiment; λ is a hyperparameter that balances classification and regression, so that the weights of the two loss functions are close. If N_{reg} is set to 256 in the code, then λ is set to 1; $L_{reg}(t_i,t_i^*)$ represents the regression loss, which is calculated by $L_{reg}(t_i,t_i^*)=R(t_i-t_i^*)$, and the function R represents the smooth L_1 loss function [6]; $p_i^* L_{reg}$ means the loss of frame regression, which is only intentional for the anchor boxes of positive samples ($p_i^*=1$), and the anchor boxes of negative samples have no practical meaning ($p_i^*=0$).

The output of the entire network is $\{p_i\}$ for the cls layer and $\{t_i\}$ for the output of the reg layer; The whole loss function adjusts the weights of the two-layer output through the three parameters of N_{cls}, N_{reg} and the balance weight λ to achieve the purpose of regression. Equation (2) is the classification loss function of cls layer, there are only two types {1,0}.

$$L_{cls}(p_i, p_i^*)= - \log[p_i^* p_i+(1- p_i^*)(1- p_i)] \quad (2)$$

Among them, the input of p_i^* is the real label of the anchor boxes, if it is a positive sample, it is 1, and the input of p_i is the probability that the anchor box with index i is predicted to be a positive sample, and the loss function of $p_i^*=1$ is $L_{cls}(p_i, p_i^*)= -\log p_i$. The closer the probability pi predicted by the classifier is to 1, the closer the calculated loss is to 0, the smaller the classification

loss. If it is a negative sample, p_i^* is 0, p_i input is the probability that the anchor boxes with index i are predicted to be negative samples, $p_i^*=0$ loss function $L_{cls}(p_i,p_i^*) =-\log(1-p_i)$. The closer the probability p_i predicted by the classifier is to 0, the closer the calculated loss is to 0, the smaller the classification loss. Through the log loss function of the two classification, the negative log likelihood function of the classifier under the condition of the real sample label can be obtained. The Equation (3) is the reg layer border regression loss function.

$$L_{reg}(t_i,t_i^*)=R(t_i-t_i^*) \tag{3}$$

Among them, t_i and t_i^* are four-dimensional vectors, calculated by formulas (4) and (5).

$$\begin{cases} t_x=(x-x_a)/w_a \\ t_y=(y-y_a)/h_a \\ t_w=\log(w/w_a) \\ t_h=\log(h/h_a) \end{cases} \tag{4}$$

$$\begin{cases} t_x^*=(x^*-x_a)/w_a \\ t_y^*=(y^*-y_a)/h_a \\ t_w^*=\log(w^*/w_a) \\ t_h^*=\log(h^*/h_a) \end{cases} \tag{5}$$

Where x, y, w, and h represent the center coordinates, width and height of the anchor boxes and ground truth bound, respectively. The variables x, x_a, and x^* are the x-coordinates of the centers of predicted boxes, anchor boxes, and ground truth bounds, respectively, and y, w, and h are similar. (x,y,w,h) represents the location information of predicted boxes, (x_a,y_a,w_a,h_a) represents the location information of anchor boxes, (x^*,y^*,w^*, h^*) represents the location information of the ground truth bound. The R in the bounding box regression by Equation (3) represents the smooth_{L1} loss function, which is insensitive to outliers and outliers, and no gradient loss occurs during training. The calculation process is shown in Equation (6).

$$\text{smooth}_{L1}(x)=\begin{cases} 0.5x^2 & \text{if } |x|<1 \\ |x|-0.5 & \text{otherwise} \end{cases} \tag{6}$$

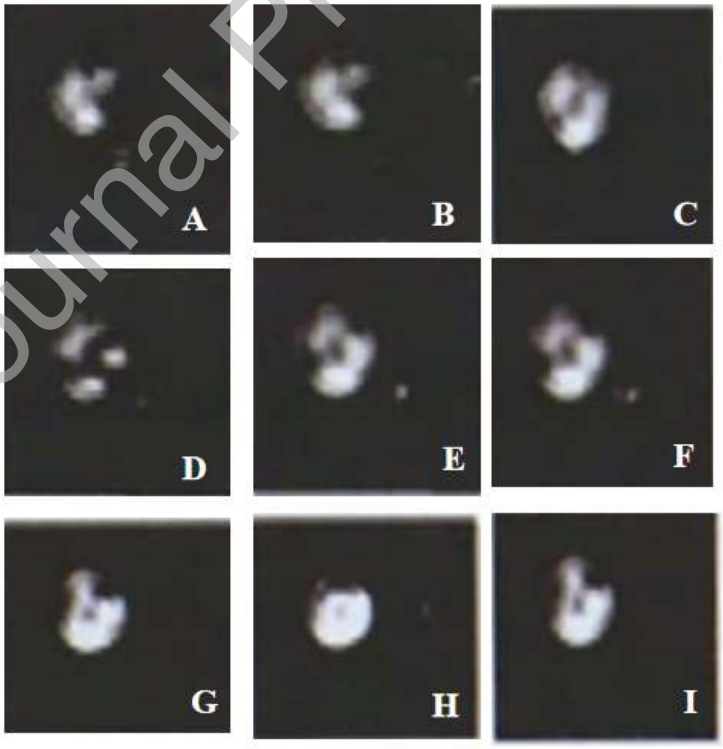
2.7 Statistical methods

SPSS22.0 statistical software was used for statistical analysis and processing. Fisher's exact probability method was used if the conditions were not met. $P<0.05$ was considered statistically significant.

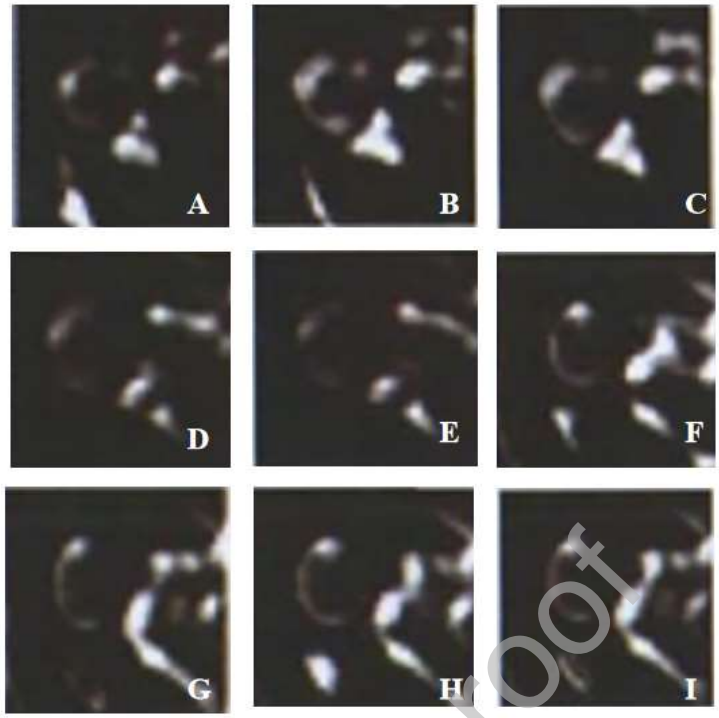
3. Experimental Results

3.1 Image segmentation result

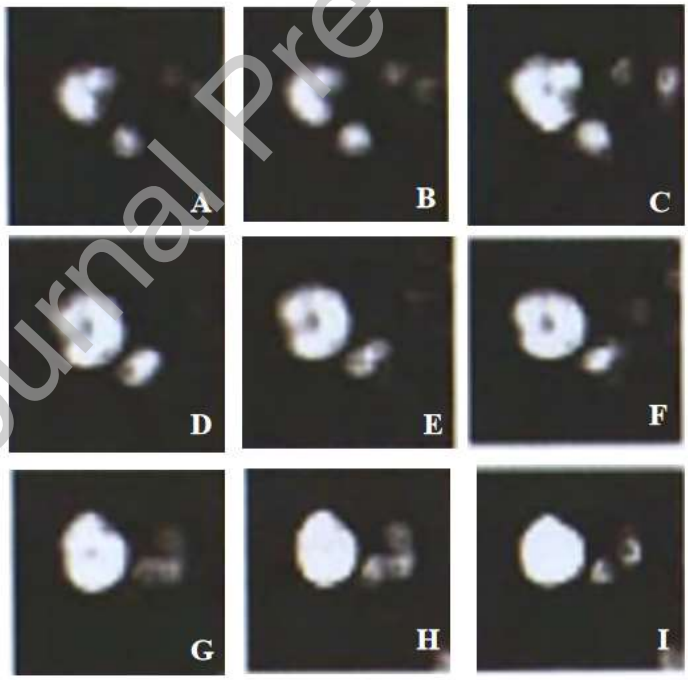
Part of the segmentation results using U-Net and AlexNet segmentation models are shown in Figure 6. Through the segmentation results of the model, it can be seen that: (1) In Figure 4(a), the articular bones show better segmentation results. This is because in the image sequence, the articular bones are simple in shape, closer to a circle or an ellipse, and the amplitude of changes between slices is small. (2) The segmentation result of the humeral head in Figure 6(b) is poor, showing more false positive areas. On the one hand, the signal strength inside the bone is very close to fat, and it is easy to confuse it with some nearby fat areas. At the same time, some fat and soft tissue regions are easily misjudged as joint bones or humeral heads in shape and global position. On the other hand, the shape of the humeral head is more complex, and the range of changes between slices is large. This variety of shapes will make it easier to misjudge noise or fat as the humeral head. (3) It can be seen from Figure 6(c) that since the positions of the two bone regions in the image are roughly fixed, they are close to the central region, and the relative positions are relatively stable, so they also show better segmentation results. (4) In Figure 6(d), AlexNet segmented the edges of all objects. Since the training uses image blocks inside or near the joint bone or humeral head, in fact, the segmentation of the joint bone or humeral head is more accurate. A-I is a schematic diagram of image segmentation in different frames of the corresponding part.



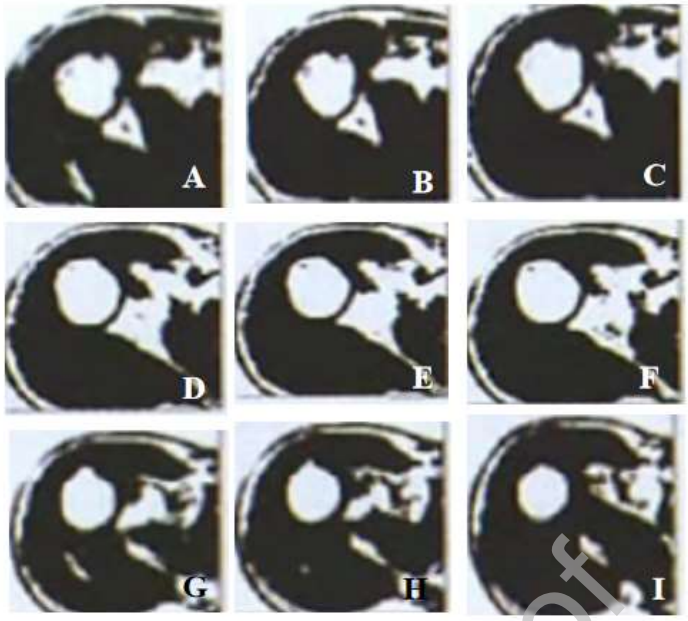
(a) U-Net articular bone



(b) U-Net Humeral Head



(c) U-Net joint bone and humeral head as a whole



(d) AlexNet

Figure 6 Model segmentation based on: (a) U-Net articular bone segmentation result; (b) U-Net humeral head segmentation result; (c) U-Net joint bone and humeral head as a whole segmentation result; (d) AlexNet segmentation result. A-I is the image segmentation in different frames of the corresponding parts.

3.2 Model training results

The experimental equipment uses a single NVIDIA Titan XGPU, CUDA-NN acceleration, VRAM is 12 GB, the operating system is Ubuntu 16.04, and the deep learning framework used is Google's Tensor-flow 1.0 platform[7]. The data set comes from 8 groups of patients, and each group of scan sequence includes about 100 images. Five groups of patients were used for training and 5-fold cross-validation, and three groups of patients were used to test the actual segmentation effect. For all patient MRI data sets, convert the DICOM image sequence to 8-bit grayscale images with a resolution of 192 × 192 pixels, and perform Gaussian filter denoising. Medical workers uniformly mark the joint bones and humeral head parts that need attention At the same time, the data set is divided into training set, validation set and test set.

For the training of the three U-Net models, the original total number of training sets for each segmentation is about 500, and the data is enhanced to about 40,000. Enhancement schemes include: random horizontal flip, random vertical flip, random offset within 10% of height and width, random 90° rotation, random zoom within 10%, random horizontal or vertical projection transformation within 10%. The point fill value of the border is 0, which is the black value. The training parameters mainly include: optimization method, batch size, iterations and epochs. The corresponding values are momentum (0.2), 1, 128, and 200, respectively. For the AlexNet model, since enough samples can be obtained, no data enhancement is performed. In the end, the training data of type 1 and type 0 are about 250,000 each. In order to prove that AlexNet has better performance in this application, AlexNet, VGG16, Inception3 and ResNet are used to evaluate the 17 × 17 pixel block. The results are shown in Table 1. It can be seen that AlexNet has the best classification performance in this application. Next, we examine the choice of the image block size. The experimental results are shown in Table 2. In the

experiment, the parameter batch size is 128, the parameter drop out is 0.5, and the number of training is 150 000 times. It can be seen that the 11 × 11 pixel block achieves the best performance, with an average accuracy rate of 98%. It is precisely because of such high-accuracy classification that the AlexNet learning model can be used to optimize the bone segmentation results based on U-Net segmentation. A-I is a schematic diagram of image segmentation in different frames of the corresponding part.

Table 1 Comparison of performance of different deep neural network models

Neural network model	Average accuracy of verification classification/%	Training speed
Alex Net	92.5	fast
VGG16	90.9	slow
Inception 3	91.1	slow
ResNet	91.5	slow

Table 2 Comparison of training and performance of image blocks of different sizes

Image block size/pixel	Average accuracy of verification classification/%
7*7	94.3
9*9	95.1
11*11	97.9
13*13	95.3
15*15	94.8
17*17	93.5
19*19	91.4
21*21	89.4
23*23	89.6
25*25	Training no longer converges

The experiment uses Dice Coefficient (D), Positive Predicted Value (V) and Sensitivity (S) to evaluate the segmentation performance, namely

$$\begin{cases} D = \dfrac{|P \cap T|}{(|P| + |T|)/2} \\ V = \dfrac{|P \cap T|}{|P|} \\ S = \dfrac{|P \cap T|}{|T|} \end{cases} \tag{7}$$

In Equation (7), P represents the voxel area of the predicted segmentation result, and T represents the voxel area of the artificially marked Ground Truth segmentation result. Table 3 shows the values of different segmentation performance metrics during the training process.

Table 3 Segmentation performance metrics

Methods	Dice	PPV	Sensitivity
Alex Net	0.91±0.02	0.95±0.03	0.95±0.02

VGG16	0.79±0.02	0.81±0.02	0.83±0.02
Inception 3	0.82±0.02	0.85±0.02	0.87±0.02
ResNet	0.80±0.02	0.79±0.02	0.83±0.02

According to the previous analysis, it is impossible to perform the correct one-time segmentation alone, so there is no feasibility for experimental comparison. Compared with other segmentation methods, it can be seen that because other segmentation methods are mainly based on the gray-scale characteristics of pixels and the existing boundary effects, they cannot learn the inherent shape and position information of the shoulder joint structure. Therefore, the accuracy of the segmentation is compromised.

The Alex Net algorithm has a high segmentation accuracy of the image, and the maximum value of Dice, PPV and sensitivity can reach 0.93, 0.98 and 0.97. It is basically consistent with the manual segmentation results, and even in a considerable part of the image (observed by partial test image segmentation results), the segmentation accuracy even exceeds other algorithms such as VGG16, Inception 3 or ResNet.

From the point of view of the experimental data set, our sample set is very limited. There are a total of 8 sets of patient use cases. As long as the image block sample method and the whole frame image data enhancement method are appropriate, very good prediction results can be achieved. Model fusion is a powerful and effective technology to enhance prediction accuracy in a variety of machine learning tasks. The voting described in this article is a method of model fusion. This idea is relatively simple, but it can greatly improve prediction performance and reduce generalization error.

4. Discussion

In the emerging stage of artificial intelligence (AI) technology, problems that computers can easily solve are relatively difficult for human intelligence[7]. The current AI technology aims to solve problems that are difficult to solve by traditional computer programs such as image recognition, speech recognition, and medical aided diagnosis, but human intelligence is easy to learn and solve. AI technology can use raw data to extract the abstract features you need to achieve the purposes of detection, positioning, and classification. This ability to extract patterns is called Machine Learning (ML)[8]. ML is the use of computers to solve real-world problems and help make seemingly subjective decisions. For example: Logistic Regression [9-10] is a simple machine learning algorithm, which can predict whether a student will pass the exam through the student's study time. In machine learning algorithms, relying on the representation of the input data (Representation), the representation is mapped to the output, that is, Representation Learning (RL) [11-12]. The autoencoder is a typical representative of learning, it is composed of encoder function (encoder) and decoder function (decoder). The input data is converted into a specific representation by the encoder function, and then restored to its original form by the decoder function. In this process, try not to lose its original characteristics. Deep Learning (DL) [13-14] uses simple representations to extract abstract, higher-level features from data and express complex representations.

Deep learning concept was proposed in 2006, originally derived from the related research of artificial neural networks, and then through continuous efforts and continuous improvement by researchers, it has developed into a branch subject with good development prospects in the field of machine learning[15]. It can be considered that deep learning is a new field in machine learning

research, a method of characterizing data based on machine learning to obtain attribute categories or features. Its motivation is to simulate the neural network of human brain analysis and learning, and obtain internal laws and theories by interpreting data (such as images, sounds and texts, etc.), and then improve and enhance its own capabilities[16-17].

Deep learning technology has now been applied in many fields, and has gradually developed from early simple applications such as static image processing and text information completion to high-dimensional applications such as video and audio data processing, from incomplete information completion to creative work[18-19]. With the support of big data, AI can quickly refine the general laws of low-level data and previous processing methods through the analysis of past data, and on this basis, digest and merge, master the processing principles, and then create new works in accordance with the principles[20-21]. Shallow creative work includes painting, writing, voice output, composition, etc., while high-level work can include intelligent driving, situation judgment, and strategy formulation[22]. Deep learning forms high-level abstract attribute categories or features by combining low-level features[23]. Therefore, data processing is the key to learning quality, and massive data processing itself is the unique strength of computers[24]. In recent years, with the development of data and network technology, the available data has increased exponentially, creating conditions for computer deep learning[25].

5. Conclusion

MRI is gradually replacing CT in the measurement and diagnosis of bone injuries and bone diseases. In this paper, for a small sample of patient data set, the voting model fusion method is used to accurately locate the shoulder and humeral head and articular bones from the four segmentation models, and the spatial coherence of the image sequence is used to predict the required area. Prediction, and then use the local perception and features of the bone region of interest to perform CNN segmentation based on image blocks, so that very accurate segmentation results can be obtained. The algorithm in this paper has been integrated into the medical image measurement and analysis platform "3DQI" developed by us, through which the 3D segmentation effect of shoulder joint bones can be displayed, and it can provide clinical diagnosis guidance to orthopedics. With the deepening of the cooperation with the hospital and the increase in the number of MRI samples, in the next step of the research, the 3-dimensional segmentation based on deep learning will be tested, and the performance indicators of the 2-dimensional segmentation will be compared and analyzed.

Conflict of Interest

The authors declare that there is no conflict of interests.

References

[1] Dalvi R, Abugharbieh R, Wilson D. Multi-contrast MR for enhanced bone imaging and segmentation[C]//Proceedings of the 29th Annual International Conference of the IEEE Engineering in Medicine and Biology Society. Lyon, France: IEEE, 2007:5620-5623.

[2] Nguyen N T, Laurendeau D, Branzan-Albu A. A new segmentation method for MRI images of the shoulder joint[C] // Proceedings of the 4th Canadian Conference on Computer and Robot Vision.

Montreal, Canada: IEEE, 2007: 329-338.

[3] Suh S W, Modi H N, Yang J H. Idiopathic scoliosis in Korean schoolchildren: a prospective screening study of over 1 million children [J]. European Spine Journal, 2011, 20(7): 1087-1094.

[4] Konieczny M R, Senyurt, Hüsseyin. Epidemiology of adolescent idiopathic scoliosis [J]. Journal of Childrens Orthopaedics, 2013, 7(1): 3.

[5] Folkesson J, Carballido-Gamio J, Eckstein F. Local bone enhancement fuzzy clustering for segmentation of MR trabecular bone images[J]. Medical Physics, 2010, 37(1) : 295-302.

[6] Folkesson J, Dam E B, Olsen O F. Segmenting articular cartilage automatically using a voxel classification approach[J]. IEEE Transactions on Medical Imaging, 2007, 26 (1) : 106-115.

[7] Schmid J, Kim J, Magnenat-Thalmann N. Robust statistical shape models for MRI bone segmentation in presence of small field of view[J]. Medical Image Analysis, 2011, 15(1): 155-168.

[8] Mallikarjuna Swamy M S, Holi M S. Knee joint articular cartilage segmentation, visualization and quantification using image processing techniques: a review[J]. International Journal of Computer Applications, 2012, 42(19): 36-43.

[9] Shelhamer E, Long J, Darrell T. Fully convolutional networks for semantic segmentation[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2017, 39 (4): 640-651.

[10] Ronneberger O, Fischer P, Brox T. U-net: convolutional networks for biomedical image segmentation[C]// Proceedings of the 18th International Conference on Medical Image Computing and Computer-Assisted Intervention. Munich, Germany: Springer，2015: 234-241.

[11] Abdulkadir A, Lienkamp S S. 3D U-Net: learning dense volumetric segmentation from sparse annotation[C]// Proceedings of the 19th International Conference on Medical Image Computing and Computer-Assisted Intervention. Athens, Greece: Springer, 2016: 424-432.

[12] Milletari F, Navab N, Ahmadi S A. V-Net: fully convolutional neural networks for volumetric medical image segmentation[C] //Proceedings of the 4th International Conference on 3D Vision. Stanford, CA , USA: IEEE, 2016: 565-571.

[13] Drozdzal M, Vorontsov E, Chartrand G. The importance of skip connections in biomedical image segmentation[C]// Proceedings of the 1st International Workshop, LABELS 2016, and 2nd International Workshop, DLMIA 2016, Held in Conjunction
with MICCAI 2016. Athens, Greece: Springer，2016: 179-187.

[14] Xie Y P, Zhang Z Z, Sapkota M. Spatial clockwork recurrent neural network for muscle perimysium segmentation[C]// Proceedings of 19th International Conference on Medical Image Computing and Computer-Assisted Intervention. Athens, Greece:
Springer, 2016: 185-193.

[15] Nie D, Wang L, Trullo R. Segmentation of craniomaxillofacial bony structures from MRI with a 3D deep-learning based cascade framework[C]// Proceedings of the 8th International Workshop. Quebec City, QC, Canada: Springer, 2017: 266-273.

[16] Andermatt S, Pezold S, Cattin P. Multi-dimensional gated recurrent units for the segmentation of biomedical 3D-data[C]// Proceedings of the 1st International Workshop，LABELS 2016，and 2nd International Workshop，DLMIA 2016，Held in Conjunction with MICCAI 2016. Athens, Greece: Springer, 2016: 142-151.

[17] Poudel R P K, Lamata P, Montana G. Recurrent fully convolutional neural networks for multi-slice MRI cardiac segmentation [C]//Proceedings of the 1st International Workshops, RAMBO 2016 and HVSMR 2016, Held in Conjunction with MICCAI 2016. Athens, Greece: Springer, 2016: 83-94.

[18] Korez R, Likar B, Pernu F. Model-based segmentation of vertebral bodies from MR images with 3D CNNs[C]// Proceedings of the 19th International Conference on Medical Image Computing and Computer-Assisted Intervention. Athens, Greece: Springer, 2016: 433-441.

[19] Zhou X R, Ito T, Takayama R. Three-dimensional CT image segmentation by combining 2D fully convolutional network with 3D majority voting[C]// Proceedings of the 1st International Workshop, LABELS 2016, and 2nd International Workshop, DL-MIA 2016, Held in Conjunction with MICCAI 2016, Athens, Greece: Springer, 2016: 111-120.

[20] Moeskops P, Wolterink J M, van der Velden B H M. Deep learning for multi-task medical image segmentation in multiple modalities[C]// Proceedings of the 19th International Conference on Medical Image Computing and Computer-Assisted Intervention. Athens, Greece: Springer, 2016: 478-486.

[21] Shakeri M, Tsogkas S, Ferrante E. Sub-cortical brain structure segmentation using F-CNN'S[C]// The 13th International Symposium on Biomedical Imaging. Prague, Czech Republic: IEEE，2016: 269-272.

[22] Song Y Y, Zhang L, Chen S P. Accurate segmentation of cervical cytoplasm and nuclei based on multiscale convolutional network and graph partitioning[J]. IEEE Transactions on Biomedical Engineering, 2015, 62(10) : 2421-2433.

[23] Christ P F, Elshaer M E A, Ettlinger F. Automatic liver and lesion segmentation in CT using cascaded fully convolutional neural networks and 3D conditional random fields[C]// Proceedings of the 19th International Conference on Medical Image Computing and Computer-Assisted Intervention. Athens, Greece: Springer, 2016: 415-423.

[24] Gao M C, Xu Z Y, Lu L. Segmentation label propagation using deep convolutional neural networks and dense conditional random field[C]// The 13th International Symposium on Biomedical Imaging. Prague, Czech Republic: IEEE，2016: 1265-1268.

[25] Alansary A, Kamnitsas K, Davidson A. Fast fully automatic segmentation of the human placenta from motion corrupted MRI [C]//Proceedings of the 19th International Conference on Medical Image Computing and Computer-Assisted Intervention. Athens，Greece: Springer, 2016: 589-597.