American Journal of Information Science and Computer Engineering
Vol. 7, No. 1, 2021, pp. 6-15
http://www.aiscience.org/journal/ajisce
ISSN: 2381-7488 (Print); ISSN: 2381-7496 (Online)

Ultrasonic Image Processing Based on DeepLab Network

Xiaotong Li, Mei Li*, Guanyi Li, Xinlin Yang

School of Information Engineering, China University of Geosciences, Beijing, China

Abstract

Because of its convenience and low price, ultrasound detection has been widely used in organ examination, especially in gynecological examination. Manual recognition and segmentation of the lesions in the image by the doctor is very heavy, and the doctor's manual interpretation of the image is easy to be affected by subjective cognition. Under a large amount of data, the efficiency is low and the error rate is high. In recent years, artificial intelligence technology, especially deep learning network, has made significant progress in medical image segmentation. It is widely used in lung cancer diagnosis and early prevention, but it is rarely used in ultrasound image processing. Most of the segmentation algorithms for medical images are based on the edge and region of the lesion. However, due to the complexity of medical ultrasound image structure, image interference noise, and changeable segmentation target, the existing algorithms can not achieve accurate lesion segmentation, so it has not been widely used in clinical, there are still many problems to be solved. DeepLab is a series of artificial neural networks, which aims at semantic segmentation task. Its network features can obtain more contextual information, and use fully connected conditional random field (CRF) to improve the ability of model to capture details. It is suitable for noise reduction and image segmentation of complex and noisy images. In this paper, combined with the deep learning neural network algorithm, the automatic segmentation of medical ultrasound image is studied and analyzed one by one. By comparing the processing effects of different deep learning networks, it shows that the deep lab network architecture has high recognition accuracy. The network can be widely used in image processing of complex lesions to improve the detection accuracy and efficiency.

Keywords

Machine Learning, Ultrasound Image, Image Segmentation, DeepLab

Received: January 29, 2021 / Accepted: March 14, 2021 / Published online: March 29, 2021

1. Introduction

With the continuous advancement of medical imaging technology and the rapid development of computer technology, medical images have become an important means for doctors in non-invasive diagnosis and treatment. Nowadays, there are many kinds of medical imaging equipment, and imaging technologies include X-ray imaging, magnetic resonance imaging (MR), ultrasound imaging, and computer tomography (CT). Medical ultrasound images contain a lot of information, which can well reflect the tissue structure of the human body. [1] Medical ultrasound imaging technology has been widely used due to its excellent performance and relatively low price.

The application of machine learning in the medical field has attracted much attention, and medical image processing is a very important application. [2]

The result of medical image processing directly affects the doctor's judgment, and image segmentation is a very important part. Traditional segmentation algorithms can no longer meet the requirements. It is necessary to learn from machine learning technology and try to use machine learning

* Corresponding author
E-mail address: maggieli@cugb.edu.cn (Mei Li)

algorithms to segment medical ultrasound images, which helps to make the images more intuitive and clear, and improve diagnosis efficiency. [3] Use machine learning methods to make further attempts on image segmentation algorithms to make up for the shortcomings of traditional algorithms.

Medical image segmentation is a key and complex step in medical image processing and analysis. The main function is to segment the parts that doctors are interested in with special meaning, and calculate important features based on the segmented image analysis. This provides a reliable basis for diagnosis and treatment. [4] The commonly used medical image segmentation methods include threshold method, region growing method, edge detection method, fuzzy clustering method, genetic algorithm-based method, wavelet transform method, neural network-based method, etc.

Medical image segmentation technology is from manual segmentation to semi-automatic segmentation through the combination of human and machine, and finally to automatic segmentation based on machine. Manual segmentation refers to a doctor with a certain experience through image editing, the image displayed on the computer, depicting the outline boundary of the lesion or tissue of interest. This segmentation is highly subjective and requires doctors to have a certain accumulation of knowledge and experience, so the group used is relatively small. With the development of computer technology, technologies such as semi-automatic segmentation algorithms have emerged. Combine computer image data processing technology with medical knowledge, and use certain algorithms to realize human-computer interactive image segmentation. Fully automatic segmentation is to iterate through a specific algorithm or model so that the computer can output the segmentation results independently and automatically. However, these algorithms are more complex and difficult to implement. The segmentation speed and segmentation results are not ideal, and there is still much room for improvement. [5]

Traditional image segmentation, due to the limited computing power of the computer, can only process some grayscale images in the early stage, and then can process rgb images. The segmentation in this period is mainly by extracting low-level features of the picture, and then segmenting, some methods have emerged: Ostu, FCM, watershed, N-Cut, etc. The result of segmentation is not semantically labeled. In other words, the segmentation does not know what it is. Subsequently, with the improvement of computing power, people began to consider obtaining semantic segmentation of images. The semantics here are currently low-level semantics, mainly referring to the types of objects that are segmented. At this stage (probably from 2010 to 2015), people consider using machines. The learning method performs image semantic segmentation. With the emergence of Full Convolutional Networks (FCN), deep learning has officially entered the field of image semantic segmentation. The semantics here still mainly refers to the types of objects that are segmented. From the segmentation results, you can clearly know what objects are segmented, such as cats, Dogs, etc. Now there is another type called instance segmentation, which can divide different objects of the same category differently, and it can be clearly known that the two people on the left and right of the segmentation are not the same person.

The craze of convolutional neural networks has led to the rapid development of the field of deep learning, and a new generation of technology appears basically every year. [6] New technologies are often accompanied by updated training methods and deeper network structures. These technologies have made significant achievements in image processing fields such as image classification, image recognition, and image segmentation.

At present, neural networks have been widely used in medical image segmentation. Scholars such as CERNAZNUGLAVAN proposed to use CNN to segment the bone contour structure in X-ray images. [7] Su, Hai proposed a fast scanning convolutional neural network, which is applied to the semantic segmentation of breast magnetic resonance images. While greatly improving the efficiency of image semantic segmentation, it also ensures that the edge segmentation accuracy of image feature parts is basically unchanged, which provides great technical support for doctors to make accurate diagnosis.

2. Selection and construction of Neural Network

2.1. Convolutional Neural Network

Semantic segmentation mainly used manual features + graph models before 2015. After 2015, a large number of solutions based on convolutional neural network (CNN) deep learning began to appear. CNN is a machine learning model under deep supervised learning. It has strong adaptability, good fault tolerance, self-learning ability and parallel processing ability. It is good at mining local features of data and extracting global training features. By combining the image's local perception area, shared weights and bias parameters, and spatially using deconvolution to make full use of the features contained in the image data itself, the network parameters are optimized, and the image displacement and invariance are guaranteed to a certain extent. [8] CNN has been successfully used to assist diagnosis and treatment in the field of medical image segmentation because of the above excellent characteristics.

Convolutional neural network is a multi-layer supervised

learning neural network. [9] The core modules for feature extraction of convolutional neural networks are the convolutional layer and the pooling layer in the hidden layer. [10] Convolution is generally used for feature extraction. Through the convolution operation, certain features of the image can be enhanced or image noise can be reduced, with the function of weight sharing and local connection.

The pooling layer is usually followed by the convolutional layer to simplify the output of the convolutional layer. The features of the local area of each feature are integrated and statistically implemented to achieve downsampling operations, such as the pooling operation that uses the maximum and average values to calculate the regional features It is called maximum pooling and average pooling, which reduces the amount of data that needs to be processed while ensuring the amount of information.

The input samples are input from the input layer, passed through layer-by-layer transfer transformation, and transmitted to the output layer, and then the corresponding actual output is calculated. After the pooling layer, an activation map can be obtained and passed to the fully connected layer. After completing a round of iteration, the parameters of each layer can be adjusted supervisedly through the backpropagation BP algorithm.

2.2. Fully Convolutional Network (FCN)

The traditional CNN-based segmentation method has large storage overhead, low computational efficiency, and the size of the convolution kernel limits the size of the sensing area, which results in the impact of classification performance. The full convolutional network (FCN) avoids the use of pixel blocks. The problem of repeated storage and convolution calculation. In 2014, the FCN network proposed by the University of California Long and others promoted and improved the original CNN network, cancelled the fully connected layer, and used deconvolution for dense prediction, so that the segmentation map was restored to the original image size.

CNN uses the fixed-length feature vector obtained by the fully connected layer after the convolutional layer to perform pixel-level classification on the image [11], while the full convolutional network (FCN) uses the deconvolutional layer to upload the last convolutional layer. Sampling operation makes it return to the same size image as the input image, so that it can accept any size input image. And FCN can generate a probability prediction evaluation for each pixel, while retaining the spatial location information in the original ultrasound image, and finally perform pixel-by-pixel classification on the up-sampled feature image. FCN converts the one-dimensional vector with the length of 4096 in the 6th and 7th layers of traditional CNN and the one-dimensional vector with the length of category N in the 8th layer into convolutional layers. All layers are convolutional layers, so they are called full convolutional networks. As shown in Figure 1.

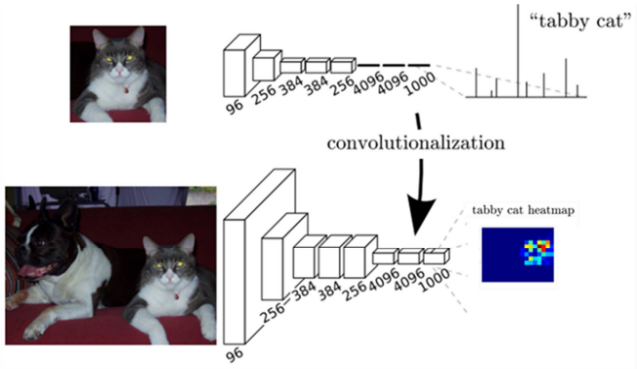

Figure 1. Full convolutional layer converted to convolutional layer.

When the output reaches the last layer, our most important high-dimensional feature map is also called heatmap. After being converted to a convolutional layer, there is no limit to the input size, but because all the convolutional layers perform dimensionality reduction operations on the image, the problem of low resolution of the output image is caused. In order to solve this problem, there is the concept of deconvolution, also called upsampling. After upsampling and zooming with the same number of layers as the convolutional layer, it is restored to the same resolution as the original image, which basically realizes semantic segmentation. The accuracy of the result obtained in the restored original image is not enough to restore the features in the image. Some details on the segmentation boundary cannot be recovered. Therefore, the output of the pooling layer of the fourth and third layers is also deconvolved, requiring 16 times and 8 times of upsampling, respectively, and then combining these results to optimize the output. Next, classify and predict each pixel, and calculate the maximum probability of the pixel's label pixel by pixel, and use it as the pixel's classification.

Although the final result has been optimized through the upsampling effect, because each pixel is classified, the connection between pixels is not fully considered, and the spatial regularization step is ignored in the pixel classification segmentation method, resulting in lack of spatial consistency. The final result is still not fine enough, it is not sensitive to the details of the feature information, and the obtained edge contour is relatively fuzzy and smooth.

2.3. DeepLab Network

Since the deconvolution in the FCN network cannot fully recover the information, semantic segmentation requires precise adjustment of the image. The traditional pooling layer

increases the receptive field of the convolution kernel and the background of the aggregated image, but also loses part of the position information. Therefore, it is necessary to retain the spatial position information discarded in the pooling layer. There are two solutions with different structures.. The combination of encoder and decoder is a solution. The encoder uses a convolutional layer and a pooling layer to gradually reduce the dimensionality of the input data, while the decoder gradually restores the details and input of the target through the convolutional layer and upsampling layer. Space dimension. Direct information connection between the encoder and the decoder can help the decoder to better restore the image details. The most typical network structure is called U-Net network. Another DeepLab method is to use the expanded convolution structure with holes to increase the field of view dimension without reducing the visual space, and remove the pooling layer structure, and finally make the network deeper and the classification result is more accurate, but because of the field of view dimension Larger leads to inaccurate positioning, so the fully connected conditional random field CRF is used for iterative optimization to refine the edge information.

The basic flow diagram of DeepLab is shown in Figure 2.

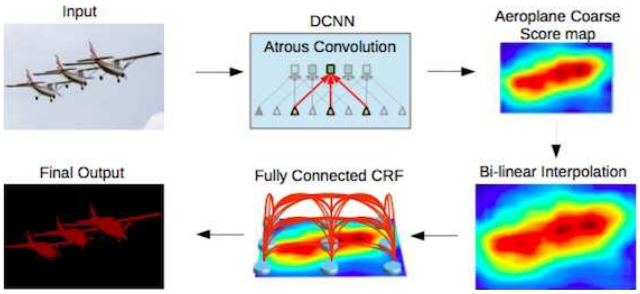

Figure 2. DeepLab flowchart.

The entire network is divided into two modules. The first module is to output rough segmentation results through DCNN, and the second module is to segment the results by fully connected CRF essence. The DeepLab architecture mainly includes ResNet structure, hollow convolution, Atrous spatial pyramid pooling, and fully connected CRF.

2.3.1. ResNet Network

The deep residual network is to solve the problem of gradient disappearance, gradient explosion and degradation. [12] It is currently the most widely used CNN feature extraction network. The ResNet network is based on the VGG19 network. When the depth of the VGG network reaches 19 layers, increasing the number of layers will cause the classification performance to gradually decrease. The basic structure of the residual network is the residual module. Figure 3 is a comparison between the ordinary CNN structure and the residual network:

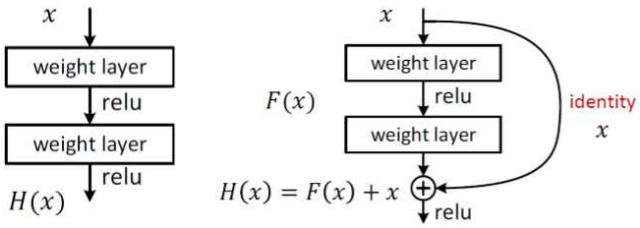

Figure 3. Common structure (left picture), residual structure (right picture).

The innovation of the residual network is to directly connect the input to the output after being weighted by two convolutions, which is equivalent to adding a shortcut between the networks or can also be called a skip connection. When the gradient is small, it can ensure that the final gradient or derivative is greater than 0 and near 1. It enables the information unit of the input node and the information unit of the output node to communicate directly. When multiple residual blocks of this structure are connected together, a residual network is formed. The training result of the residual network will not deteriorate when the network continues to deepen, which can solve the problem of network degradation and avoid overfitting to a certain extent. In addition, the residual network cancels the maximum pooling and fully connected layer, which can simplify the network and avoid the limitation of the size of the input ultrasound image.

2.3.2. Hole Convolution

In the field of image segmentation, traditional convolutional networks, such as FCN, first convolve the input image and then pool it. After the pooling operation is performed, the image size and spatial dimension are reduced while increasing the corresponding receptive field of convolution.[13] However, the final predicted segmentation result of the image requires a clear pixel category output, so the pooled low-pixel image is subjected to deconvolution and upsampling to restore it to an image of the same size as the original image for inference. Since the pixel size on the feature layer after the image is pooled is relatively low, the accuracy and spatial hierarchical information of the image will be lost even by upsampling, resulting in the inability to reconstruct small object information after deconvolution, and the performance cannot be significantly improved. Therefore, the hole convolution abandons the traditional pooling layer to make the receptive field of each layer of the network smaller. Although it reduces the prediction accuracy of the entire model, it avoids the loss of image feature information due to the loss of some pixel information in the downsampling operation.

Hole convolution or dilation convolution, the main idea is to introduce a new parameter called "dilation rate" to the traditional convolution layer, which defines the value of each pixel when the convolution kernel processes data. Spacing, while removing the pooling operation, does not reduce the

receptive field of the network, thereby ensuring the accuracy of semantic segmentation of ultrasound images. The principle is shown in Figures 4 and 5.

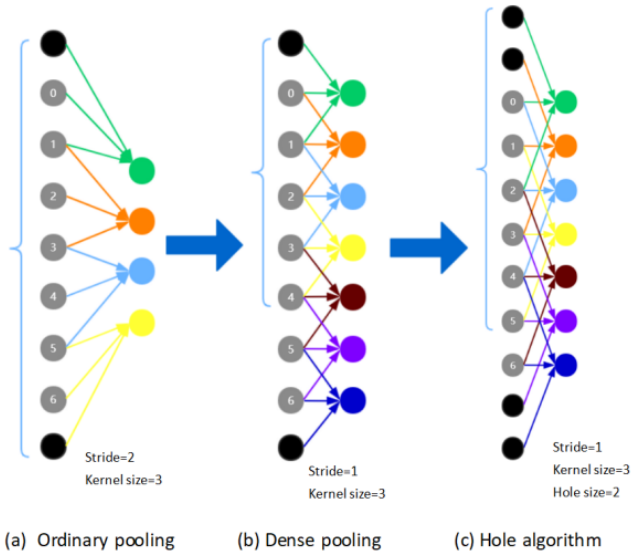

Figure 4. Comparison of pooling and hole convolution.

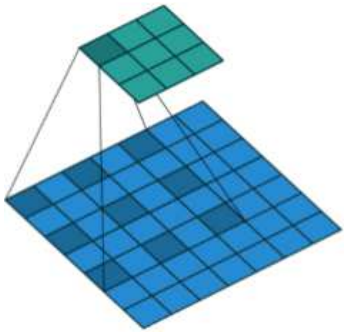

Figure 5. Principle of Hole Convolution.

2.3.3. Atrous Spatial Pyramid Pooling

In an ordinary CNN structure, a fixed-size image is usually input for training and testing. For images with different sizes, if they are converted to a uniform size through operations such as scaling and cropping, it may cause missing and deformed image information, invisibly increasing the weight of certain regions, thereby reducing the accuracy of recognition and detection.

In order to improve the computational efficiency of the algorithm, the spatial pyramid pooling uses multiple sampling rates to perform convolution in a specific feature layer, and uses multiple complementary and effective convolution kernels to detect the input ultrasound image.

Therefore, different convolution kernels can be used in effective images to capture characteristic objects at multiple scales. Compared with the traditional pooling layer for re-sampling features, this parallel porous mapping is performed on the convolutional layer with different sampling rates and different receptive fields, and a new type is added between the convolutional layer and the fully connected layer. The network layer named Atrous spatial pyramid pooling, while allowing the fully connected layer to output a fixed number of features, in this way, training can be carried out without loss of image location information and as little feature changes as possible. After the convolutional layer, each image is extracted with different sizes of receptive fields, so that the feature information of the original image is retained to the greatest extent.

For example, when we input an image of any size and want to extract 21 features, we first carry out 4 parallel void convolutions with rates of 6, 12, 18 and 24 respectively, and the size of the convolution kernel is all 33. When the convolution of four voids acts on a pixel of the input image, the range of receptive field is 1313, 2525, 3737, 4949 respectively. Then add two volumes convolved with a kernel size of 11, as shown in Figure 6.

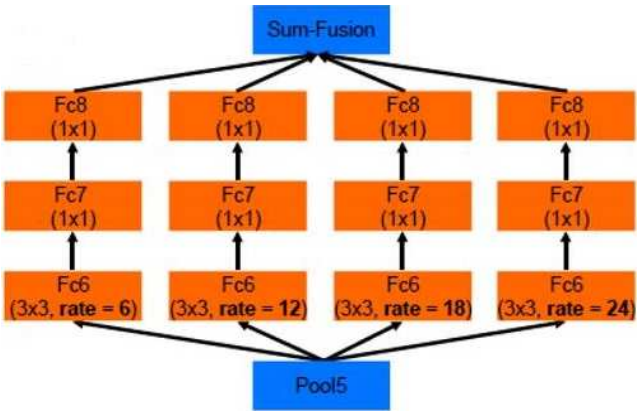

Figure 6. Pixel feature inference process.

In the 6th layer, the output size of the hole convolution is 4096, and the output of the 7th layer is 1024. Finally, the decision output is made through the 8th layer of convolution, and the 21 feature categories we need are output. Each pixel passes through the hole convolution of 4 different receptive fields to obtain the probabilities of 21 categories, Finally, the pixel-by-pixel probability addition operation is performed, and the highest probability among the 21 categories decided by the 4 branches is calculated as the final inference result. That is, the receptive fields of different sizes brought by the hole convolution of different rates are used to capture the features on different scales, thereby bringing about a more stable image segmentation effect.

2.3.4. Fully Connected CRF Essence Segmentation Result

There is usually an unavoidable drawback in the DCNN network, that is, the deeper the number of layers of the network model, the better the classification effect, but as the

number of layers increases, the larger the receptive field will result in smoother results and blurred location information. Although the hole convolution increases the resolution by 4 times, it is still not fine enough for the original image resolution. The reason is that the segmentation results produced by deep learning neural networks are often relatively smooth, and the classification results produced by ordinary CRF weak classifiers are not satisfactory. The model structure will be coupled with the nodes adjacent to the edge of the image feature part, and the feature points of the same mark will be assigned to the pixels that are close in space. Although the image feature positioning can be enhanced, small structures are still missed, making the local structure smoother, so it can only be used for smoothing and denoising.

The fully connected conditional random field CRF is used to improve the ability of the model to capture details and enhance the classification ability at the pixel level. When the CRF is coupled with the deep convolutional neural network, better results can be obtained. The energy function used by the fully connected CRF model is: $E(x) = \sum_i \theta_i(x_i) + \sum_{ij} \theta_{ij}(x_i, x_j)$.

Where x is the label assigned by the pixel, the first half is the energy value of the node itself, and the second half is the energy value of the relationship between the nodes.

Output the feature probability of the Atrous space pyramid pooling, and calculate its single point potential energy by $\theta_i(x_i) = -\log p(x_i)$. The paired potential energies have the same form of effective inference, connecting all the pixels of the image through $\theta_{ij}(x_i, y_j) = \mu(x_i, y_j)\left[\omega_1 exp\left(-\frac{\|p_i - p_j\|^2}{2\sigma_\alpha^2} - \frac{\|I_i - I_j\|^2}{2\sigma_\beta^2}\right) + \omega_2 exp\left(-\frac{\|p_i - p_j\|^2}{2\sigma_\gamma^2}\right)\right]$. $\mu(x_i, x_j)$ represents the positional relationship of pixels, which is 1 when i is not equal to j, and 0 otherwise. $\|p_i - p_j\|$ represents the distance in the probability space, and $\|I_i - I_j\|$ represents the Euclidean distance between two pixel nodes in the color space. Others are hyperparameters to control the scale of the Gaussian kernel. The first kernel emphasizes that pixels with similar colors and positions have similar categories, while the second kernel only emphasizes the spatial similarity when the image is smoothed. This model can be analogous to probability inference. The Gaussian filtering algorithm significantly accelerates the entire calculation process and improves the performance of the entire deep neural network.

3. Experiments and Discussion

There are many tasks in computer vision, including image classification, target detection, semantic segmentation, instance segmentation and panoramic segmentation. This article is mainly to achieve semantic segmentation to segment the lesion from the ultrasound image. Semantic segmentation is mainly classified at the pixel level, and each pixel is labeled, and the same label is divided into a category.

3.1. Install the Experimental Environment

It adopts the development environment of windows + python + tensorflow + gpu, specifically it runs on the python3.6 +tensorflow1.10 + CUDA9.0 platform and environment. TensorFlow is the second-generation artificial intelligence learning system developed by Google. It can be used in multiple machine deep learning fields such as speech or image recognition, and uses data flow graphs for numerical calculations. And it supports various operations such as setting the hole expansion degree of convolutional neural network and pooling upsampling expansion degree, which is very friendly to the network structure proposed in this paper [5]. It can run on multiple CPUs and GPUs. Since deep learning algorithms need to run on large data sets, it is very beneficial to run these algorithms on CUDA-enabled Nvidia GPUs to achieve faster execution speeds. At the same time, the acceleration package cudnn corresponding to CUDA needs to be installed.

The versions of Tensorflow, python, CUDA and cudnn need to be installed together. Anaconda will be installed before installing python, because it can help us easily manage and specify the separate environment of the Python distribution without interfering with the python version installed on the system. First create a new environment according to your own needs, enter conda create -n tensorflow-gpu python=3.6 to install Tensorflow in the new environment, in order to improve the training speed, you need to install the gpu version pip install tensorflow-gpu==1.10, run in python

sess = tf.Session (config = tf.ConfigProto (log_device_placement = True)) command and check if it recognizes the GPU. As shown in Figure 7.

```
>>> sess = tf.Session(config=tf.ConfigProto(log_device_placement=True))
2019-05-11 21:45:56.825646: I T:\src\github\tensorflow\tensorflow\core\platform\cpu_feature_
2019-05-11 21:45:57.212977: I T:\src\github\tensorflow\tensorflow\core\common_runtime\gpu\gp
name: GeForce GTX 970M major: 5 minor: 2 memoryClockRate(GHz): 1.038
pciBusID: 0000:01:00.0
totalMemory: 3.00GiB freeMemory: 2.48GiB
```

Figure 7. Installation result.

3.2. Image Data Preprocessing

3.2.1. Image Filtering and Denoising

Ultrasound is a sound wave with a frequency greater than 20KHZ that the human ear cannot feel. It can propagate in various media. It is a longitudinal wave or a transverse wave, and it has the same physical properties as sound waves [1]. Features such as the Puller effect. In addition, the ultrasonic frequency is high, the wavelength is short, and it also has strong penetration and good directivity. When the ultrasound probe scans the surface of the human body in a certain direction, the ultrasound will propagate in the human tissue. When it encounters a lesion tissue that is different from the normal tissue, acoustic impedance will be generated, and the sound wave will be scattered and reflected on the surface of the tissue, resulting in echo Wave signal. After receiving the echo, the ultrasonic equipment will undergo a certain conversion process, and then use signal processing technology to finally display the information as an image. Since the echo intensities of normal tissues and diseased tissues are different, pathology and medical knowledge can be analyzed and summarized to make certain inferences about the location of the lesion.

Medical ultrasound images have equipment information at the bottom and edges, and occupy a large pixel ratio. Clipping these equipment information can greatly increase the difficulty of data set production and the speed of network training. When the ultrasound propagates in the human body, it will produce refracted waves on the surface of various internal organs and interfere with the surrounding scattered waves. In addition, the patient will inevitably have breathing movement during ultrasound imaging. Finally, the equipment will receive the echo signal and display it on the image. Appears as speckle noise and a lot of artifacts. These noises and artifacts will affect the results of segmentation detection, so it is very necessary to denoise the image. The specific image preprocessing process is shown in Figure 8.

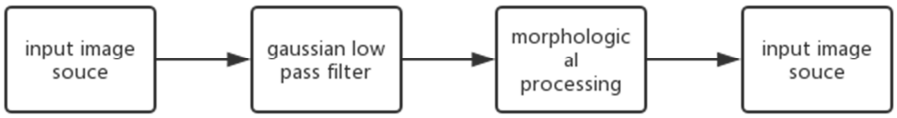

Figure 8. Flow chart of ultrasound image preprocessing.

In the field of image processing, a lot of noise is similar to Gaussian noise, because the normal distribution is widespread in many places in life, using Gaussian low-pass filtering with a standard deviation of 2 can effectively make the ultrasound image smoother, and use morphology to delete small-area objects. The processing eliminates the small area speckle and then uses the open operation to eliminate the small protrusions and make the outline smoother.

3.2.2. Extract Features and Label

Semantic segmentation is to understand the image at the pixel level, so we have to mark each pixel as a specific category. This topic only needs to segment the pathological part, so there are only two categories of background and adenomyoma. Need to circle the correct lesion site in the ultrasound image. The following is a schematic diagram of using labelme to make a semantic segmentation data set, as shown in Figure 9:

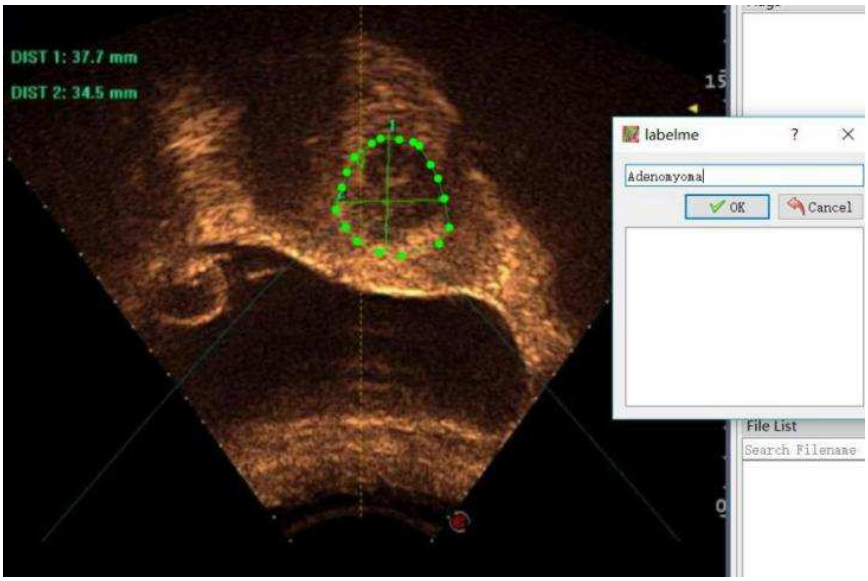

Figure 9. Schematic diagram.

After saving, convert the generated json file into a label image, and finally generate a binary image through a simple python script, with the background pixel being 0 and the labeling area being 1.

3.2.3. Create Training Set and Test Set

The data sample of the training set is an important factor affecting the performance of the classifier. In the machine learning neural network, a large amount of data is needed to provide rich feature information for the training of the classifier, and enough data samples can guarantee the accuracy of the classifier. In addition, the label information of the training set must be true and reliable to make the actual application of the classifier more accurate and reliable.

The data samples used in this article are uterine ultrasound scan images and videos provided by doctors as experimental data, which have high authenticity. The image obtained from the ultrasound equipment contains personal information and ultrasound equipment information, and the effective area is relatively small. It is necessary to simply crop the image data to obtain the effective area and remove the text area in the image. The removal of patient and equipment-related information is conducive to protecting the privacy of patients, and is more conducive to training and improving the accuracy of image segmentation. This subject uses 513 ultrasound images acquired as a sample set. If there are too many training samples, this will easily lead to too long training time, but if there are too few training samples, it will cause the classifier to be in an "under-learning" state, which greatly affects the final segmentation result. In order to obtain an appropriate number of data samples, 513 images were flipped horizontally, so that a total of 1026 image sample sets were obtained. Among them, 684 image samples are used as the sample set of the training algorithm, and 342 images are used as the sample set of the test algorithm. The training sample set is used to adjust the weight and bias of the network during the training phase, and the test set is used to test the performance of the network model during the training process, and adjust the requirements of the network model or increase the number of training cycles. Since this article uses the Tensor Flow platform, it is necessary to convert the final training data sample and test data sample into a TFRecord format that can be understood by Tensor Flow.

3.3. Training Network Model

The parameters of the model should be initialized before training. The image samples in this article are relatively small. In this case, it is difficult to retrain the parameters in the network layer and it is difficult to adjust. Therefore, the fine-tune method is based on the trained model. Use your own image samples to continue training, and fine-tune the previous parameters according to the characteristics of your own data. So you can avoid the problem of reinitializing the parameter model and failing to obtain convergence, thereby improving the efficiency of deep learning network training.

3.4. Test Results

The test results are shown in Figure 10.

Figure 10. Training result display.

The left image is the original input image, the middle image is the segmentation result generated by the trained model, and the right image is the coupled image generated by image fusion. The final result of the accuracy of testing the entire test sample is shown in Figure 11.

```
at there could be performance gains if more memory were available.
2019-05-12 15:03:44.234699: W T:\src\github\tensorflow\tensorflow\c
ntime\bfc_allocator.cc:219] Allocator (GPU_0_bfc) ran out of memory
locate 2.57GiB. The caller indicates that this is not a failure, bu
at there could be performance gains if more memory were available.
Intersection over Union for each class:
    class 0: 0.9447
    class 1: 0.2723
mean Intersection over Union: 0.6085
Pixel Accuracy: 0.9458
```

Figure 11. Evaluation of training results.

Pixel Accuracy is the most intuitive measurement method. It directly calculates the ratio of the number of pixels correctly classified to the total number of pixels, and the segmentation results of all test samples can also be output through the trained model as shown in Figure 12.

Figure 12. Output the training results of the test set.

3.5. Result Analysis

The final result of the experiment can basically meet the expected goal, the position of the lesion can be accurately located, and the effect of the boundary segmentation result needs to be improved. There are two important reasons: 1. All images are labeled by me. My lack of medical knowledge may make the boundary labeling of the data sample itself inaccurate. 2. Too few ultrasound image data samples may cause the neural network model to overfit. Due to time and computer performance issues, the number of training sessions is small, and the training results may not converge strictly, resulting in a small number of unclear segmentation contours.

4. Conclusion

The application of deep learning neural networks in medical image segmentation has developed rapidly in recent years [14]. More and more people realize that neural networks have important theoretical significance and practical application value in the field of medical segmentation processing, and the architecture of neural networks is rapidly updated. Compared with CNN and FCN, DeepLab has higher recognition accuracy [15]. The segmentation results of adenomyoma images show that the effect is consistent with expectations.

The training of deep learning neural networks usually requires a large number of data sets to support. However, because of the sensitivity and particularity of medical images, image annotation requires a large number of medical experts and scholars to complete manually, which is efficient and low-cost. So semi-supervised or unsupervised learning methods can be used to overcome the problem of lack of data or unavailability of data without affecting the accuracy of the medical system. This is a direction that urgently needs further research. It is hoped that a medical image library can be established in the future to share data resources for different medical service providers and provide more data support for artificial intelligence researchers. After the algorithm is relatively mature, it is combined with embedded and applied to computer-aided diagnosis equipment, thereby improving the efficiency and accuracy of diagnosis.

Due to limited time and level, there are some other excellent algorithms, the method proposed in this article still has much room for improvement and optimization. The algorithm in this paper does not consider the training time issue in the results. In the case of too long training time, the deep learning network model has a large number of hidden layers. To obtain high-level features from low-level features and back-propagation to optimize model weights requires a lot of calculations, so we can use the parallel framework for training.

Acknowledgments

This paper is supported by State Key Laboratory of Computer Architecture (ICT, CAS) under Grant No. CARCH 201806.

References

[1] Cheng Yiping.. Denoising algorithm for medical ultrasound images based on shear wave transform based on translation invariance. (Doctoral dissertation).

[2] Yang Yuanhang. (2018). Deep learning-oriented medical image analysis system and its practice in gastroscopy video segmentation. (Doctoral dissertation).

[3] Wang Wei. Research on image quality evaluation methods based on machine learning. (Doctoral dissertation).

[4] Zhang Lei, Zhang Minghui, Lu Zhentai, Feng Qianjin, & Chen Wufan. (2015). Brain image segmentation based on multi-weighted probability map. Journal of Southern Medical University, 35 (008), 1143-1148.

[5] Huang Yang. (2016). Variational level set image segmentation based on entropy. (Doctoral dissertation).

[6] Lu Enhui. (2019). Research on image classification based on convolutional neural network. (Doctoral dissertation).

[7] Tu Yongcheng. (2019). Research on computer-aided measurement algorithm of Cobb angle in scoliosis images. (Doctoral dissertation).

[8] Shi Dongli, Li Qiang, & Guan Xin. (2018). Brain tumor segmentation combined with convolutional neural network and fuzzy system. Computer Science and Exploration, 012 (004), 608-617.

[9] Rui Ting, Fei Jianchao, Zhou You, Fang Husheng, & Zhu Jingwei. (2016). Pedestrian detection based on deep convolutional neural networks. Computer Engineering and Applications, 52 (013), 162-166.

[10] Lv Lijing. (2016). Gland segmentation in colon pathological images based on convolutional neural network. (Doctoral dissertation).

[11] Yang Jianfeng, Hao Xiangyang, Ye Yu, Li Pengyue, & Zheng Kai. (2020). Research on cloud and snow detection method based on deepLab v3+ based on multispectral image of Tianhui-1 satellite. Science and Engineering of Surveying and Mapping (4), 40-45.

[12] B. A. Lauze. (2014). Ultrasound image segmentation.

[13] Qiang Kunkun, Song Qingyun, Luo Hong, Yang Taizhu, Wen Juan, & Zhang Wen et al. (2019). Ultrasound image characteristics and clinicopathological analysis of ovarian goiter. Western Medicine, 31 (04), 98-103.

[14] Zhao Xia, & Ni Yingting. (2020). Object part segmentation network based on deepLab. Pattern recognition and artificial intelligence, v. 33; No. 201 (03), 24-33.

[15] Xue Fei, Wu Yueqing, Yao Yu, & Ren Wei. (2019). Left ventricular ultrasound image segmentation method based on deepLab v3. Computer application.